# Enterprise Architecture Traces on the Web*

An Ontology-driven Integrative Review

Nico Gießmann

*Institute for Multimedia and Interactive Systems, University of Lübeck, Germany*

## Abstract

This paper introduces *enterprise architecture (EA) traces*, a new theoretical lens for collecting EA artifacts from web sources to accelerate enterprise modeling. EA traces are defined as data publicly accessible on the web, providing insights into an enterprise's architecture, often without the originator's awareness, offering a rich, yet complex interpretive device for understanding the peculiarities and evolution of EA. This paper first conceptualizes EA traces, then performs an integrative review encompassing 119 eligible records. The objective of this review was to identify which enterprise-related insights have been previously derived from web data, and whether they can be mapped to EA artifacts. The results reveal promising coverage across architecture layers and substantial potential for extension. Despite this, the approach lacks methodological guidance and remains highly subjective. Therefore, a three-step method is proposed drawing on ontology mapping. Utilizing selected studies from the review, the feasibility of this method is demonstrated by deriving employees' competencies from LinkedIn profiles. The paper concludes with a research agenda to guide future work on EA traces.

## Keywords

Enterprise architecture, Enterprise modeling, Digital trace data, Corporate disclosure, Web-based data collection, Ontology mapping

## 1. Introduction

Manual collection of EA artifacts is both time-consuming and prone to errors [1], which makes it one of the biggest challenges for EA management in practice [2]. Automated approaches are being discussed as an initial step in the modeling pipeline to ease this burden. The majority of automated approaches depend on internal data sources [3], which present challenges related to data availability and integration, thereby diminishing the advantages inherent in automated approaches.

The sourcing of enterprise-related web data is therefore proposed, utilizing for example the semantic web [4] or machine learning [5]. The web facilitates the dissemination of this type of data [6]. Individuals disseminate data about their careers, share insights on their employers, and provide feedback through online platforms [7]. Moreover, companies disseminate data about themselves to comply with regulatory mandates, or to present themselves to prospective customers, for instance, on their corporate websites [28], in job advertisements [29, 30], or on social media [31].

Research on web-based data collection in the EA domain is sparse [32]. In one of the few studies in the field, [32] utilized natural language processing to identify business activities from corporate websites' HTML tags. This enabled the formulation of abstract representations of the enterprise's business processes and the supporting technical infrastructure. This status quo motivates investigating the integration of enterprise-related web data into EA modeling practices. This could facilitate the acceleration of EA modeling and offer researchers in the information systems discipline a novel, substantial data source. Yet, there is a lack of both theoretical and methodological guidance [5]. Consequently, this paper provides a conceptualization of the topic as an integrative review [8], and a novel method is proposed.

*PoEM2025: Companion Proceedings of the 18th IFIP Working Conference on the Practice of Enterprise Modeling: PoEM Forum, Doctoral Consortium, Business Case and Tool Forum, Workshops, December 3-5, 2025, Geneva, Switzerland*

* References starting with No. 28 are records identified in the literature search and can be found in the online appendix: https://osf.io/692fy

nico.giessmann@uni-luebeck.de (N. Gießmann)
https://www.imis.uni-luebeck.de/de/institut/team/nico-giessmann (N. Gießmann)
0009-0007-4712-2338 (N. Gießmann)

To build such a conceptualization, it is essential to consider how enterprise-related data becomes available on the web. The *data-generating process* (DGP) refers to the set of social, technical, and methodological mechanisms through which observed data about social phenomena are produced [9]. Looking at the DGP that bring forth the sought-after data, two distinct concepts emerge in research:

1. *Digital trace data*, also referred to as digital footprints [10] or digital exhaust [11], is defined as "records of activity (trace data) undertaken through an online information system (thus, digital)" [12, p. 769]. One research stream considers them to be stored in private IT systems, such as log files [13]. Others suggest that at least some traces are public [14]. In essence, users are unaware of the digital traces they leave behind [15]. Researchers have pointed out that when an individual becomes aware of their digital traces being utilized, this can lead to manipulation of traces, thereby raising concerns about the validity of the results [12].
2. For the second DGP, "firms have incentives to communicate information to their various stakeholders, including investors, suppliers, customers, and employees" [16, p. 1]. This deliberate, and often event-based form of communication can be classified under the overarching concept of *corporate disclosure*. It encompasses two categories of information. The first includes mandatory information that is nowadays provided via the web (e.g., earnings announcements). The second category includes information that serve a purposeful communicative function to the firm (e.g., marketing or job advertisements). Corporate disclosure is indicative of the manner in which enterprises want to be perceived. These perceptions are subject to manipulation.

Digital traces and corporate disclosure are both crucial to capturing a holistic perspective of the enterprise. Corporate disclosure tells an intentional narrative of how the enterprise (and its stakeholders) want to be perceived (e.g., product claims on corporate websites). In contrast, the low awareness of digital trace data can reveal the actual behavior, or how the enterprise operates in practice (e.g., users report long wait times in reviews). The example shows that, in our context, awareness of digital traces must be re-framed in such a way that although individuals knowingly publish enterprise-related data online, they are not aware that this data can be used to draw conclusions about the enterprise's internal practices. Previous studies on enterprise-related web data [5, 17][32] have largely assumed intentionality, thereby overlooking the critical insights trace data can offer.

I therefore propose a new concept, *EA traces*, for this novel interpretation of enterprise-related web data. I posit that both EA artifacts and enterprise-related web data stem from distinct domains, ontologically speaking, which cannot be mapped without interpretation. From the differing DGP, it becomes evident that a method to map enterprise-related web data to EA artifacts needs to account for biases when interpreted, i.e., platform constraints and algorithmic confounding for digital trace data and signaling for corporate disclosure. Therefore, it is inherent that EA traces cannot be understood as either EA artifacts or enterprise-related web data, but instead as an interpretive device between the two. Following this, I have derived a definition of EA traces: *EA traces are data publicly accessible on the web, providing insights into an enterprise's architecture, often without the originator's awareness, offering a rich, yet complex interpretation device for understanding the peculiarities and evolution of EA.*

This paper is organized into four sections. Section 2 performs an integrative review of enterprise-related web data investigated by previous research. Section 3 then proposes a method to map enterprise-related web data to EA artifacts, thereby conceptualizing the interpretive device. This is followed by a demonstration of the three-step method using LinkedIn profiles and competency artifacts in section 4. Last, section 5 concludes with a research agenda to encourage further investigation into EA traces.

# 2. Integrative Review

## 2.1. Method

An integrative review was conducted to identify and synthesize research across multiple domains that provides direct or latent evidence for EA traces [18]. The scope of this review can be defined along the taxonomy proposed by [19] (see Table 1). The literature search was conducted through a targeted selection of keywords. The databases yielded no results for narrow keywords ("enterprise architecture AND (traces OR disclosure)"). Therefore, the keywords comprise a part of the 5W1H format covering the originators of enterprise-related insights (Who), the way of distribution (as verbs) or the types of web data available (as nouns) (How), and the sources (Where). The online appendix presents a full table of the keywords' logic. The keywords were transformed into a search string consisting of the four groups connected by "AND" and within-terms connected by "OR." Wildcards were added to the terms.

**Table 1**
Scope of the review.

| Characteristic | Categories | | | |
|---|---|---|---|---|
| Focus | **Research Outcomes** | Research Methods | Theories | Practices or Applications |
| Goal | **Integration** | Criticism | Identification of Central Issues | |
| Perspective | Neutral Representation | **Espousal of Position** | | |
| Coverage | Exhaustive | Exhaustive with Selective Citation | **Representative** | Central or Pivotal |
| Organization | Historical | **Conceptual** | Methodological | |
| Audience | **Specialized Scholars** | General Scholars | Practitioners or Policy Makers | General Public |

## 2.2. Literature Search

The records were collected and screened according to [20]. To cover a diverse range of outlets, I used four databases: Web of Science, Scopus, Springer Link, and the ACM Digital Library.[1] The filters applied directly when sourcing included title search and the exclusion of abstracts, preprints, retractions, and grey literature. The language was set to English and German. Moreover, the year was set to exceed 2000, as this marked the apex of the dot-com bubble and the subsequent establishment of the web economy. The subsequent processing steps are illustrated using the PRISMA flowchart (see Figure 1).

The records were consolidated and processed automatically. This step involved removing duplicates, retracted records (if still present), and predatory or vanity publishers, as per Beall's list. I focused on including high-quality research (SJR Q1-Q2). Outlets not included in the ranking were manually examined to ensure that only sources that underwent peer review were included.

The full text was assessed for eligibility, with the criteria depicted in the online appendix. Records that did not provide enterprise-related insights, such as studies about online discourse or marketing, were excluded. Studies without empirical investigation (e.g., method papers, reviews) were also excluded since they do not provide unique data sources for enterprise-related insights. Private or artificial data, including data behind a paywall, server logs, and mock-ups for experiments, were also excluded. Records that solely investigated enterprise-related web data descriptively without providing insights or based on statistical inference are not in the scope of this review. The public and nonprofit sectors were excluded due to the significant discrepancy to firm data, primarily because of higher disclosure obligations (e.g., the Freedom of Information Act).

## 2.3. Results

### 2.3.1. Synthesis

The SLR identified 119 records from various academic disciplines during the years 2001-2025 that gathered enterprise-related insights from web sources. The results are organized according to the 5W1H format previously employed: originator (Who), content type (How [does this insight become public?]), web source (Where), and enterprise-related insight (What). This organization is illustrated in table 3.

[1]The database extraction was performed on May 21, 2025.

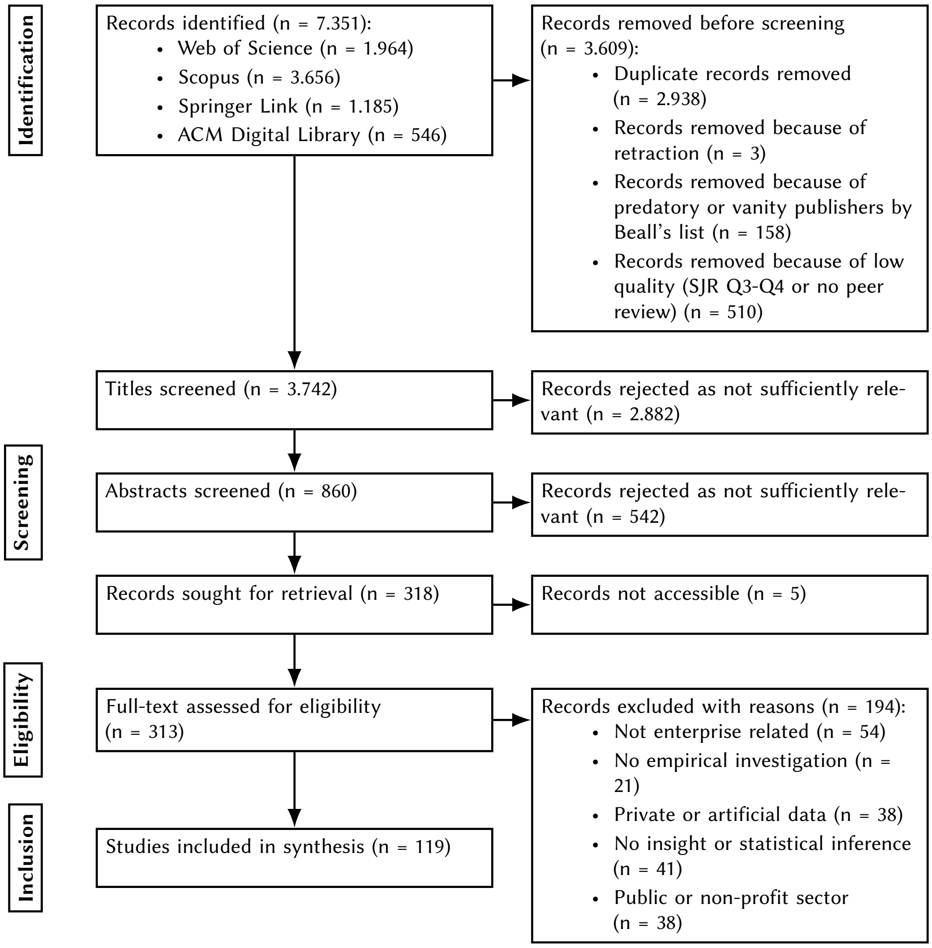


**Figure 1:** PRISMA flowchart.

The review yielded a total of eight unique *originators*. These largely align with the differentiation between firm-generated and user-generated content within the marketing domain. User-generated content is further differentiated into (retail) investor, customer, and employee categories. The government, press, and third-party data aggregators constituted the other originators. Platforms can also originate enterprise-related insights, such as Wish.com providing sales statistics of products [33].

The most frequently utilized *content type* are social media posts. A further differentiation was made between technical app reviews on the Apple App Store or Google Play Store and reviews on forums (e.g., TripAdvisor, Yelp, Glassdoor). A variety of firm-generated content exists, including social media posts, press releases, patents, and job advertisements. The government plays a dual role in creating and disseminating content. For instance, it conducts national surveys, e.g., the Mexican Social Security Institute [34]. Additionally, it functions as a provider of public registries, e.g., the German Commercial Registry [35]. The press disseminates enterprise-related content via two channels: news articles published on their websites, as seen in The Economic Times [36], and via news aggregators, such as the Factiva service by Dow Jones [37] or investing.com [38].

The most frequently utilized web sources are social media platforms. These platforms offer researchers a streamlined approach to data collection, facilitating the integration and processing of large volumes of data in a consistent manner, as opposed to the more labor-intensive process of scraping data from individual company websites. Endeavors have been undertaken to utilize the Wayback Machine to facilitate this process [28]. A similar platform dynamic can be observed for the Apple App Store and Google Play Store in the sourcing of customer app reviews.

This content was used to gain *insights* into the enterprise. I refrain from naming these EA artifacts at this point, as except [32], no research has situated their work in the EA domain. From an outside perspective, strategic insights were gathered from a market platform [39] and recalls from a government database [40]. Social media posts were used to derive (retail) investor sentiment [41] and a business's downtime during crisis [42]. In addition to financial data, the review identified studies into sustainability [28], circular economy actions [43], and corporate social responsibility [44]. Further, the stakeholder perspective was investigated, encompassing business ecosystems [35, 45, 46] and supply chain risks [47]. An additional category encompasses products of a company and how customers perceive them. The product's features were deduced from the company website [48, 49]. Its similarity to competing products was determined through user forum posts [50, 51]. The subsequent investigation of customer satisfaction and service/feature requests was conducted through various channels, including social media posts [52], app reviews [53], and national survey data (e.g., from healthcare providers [34]). This external perspective is further investigated by inferring the marketing tactics via firms' press releases [37], their brand personality [54, 55], and content virality [56]. From an internal perspective, business processes were identified from the company website [32], and its business model innovations were investigated via social media [57, 58]. [59] investigated the organizational culture, with a specific focus on HR practices [60], including employees' competencies [61] and job-related challenges [62].

#### 2.3.2. Coverage of EA Artifacts

To demonstrate how these enterprise-related insights can be understood as EA artifacts, a mapping to the layers of EA was conducted. [21] proposes a five-layer model of EA, which will be employed using the ArchiMate 3.2 framework. In this paper, the framework solely serves the purpose of providing a visual representation. Figure 2 illustrates the ArchiMate 3.2 framework, with enterprise-related insights from the review mapped to their corresponding EA artifact. The assignment was done following the official specification and was guided by the examples of practitioners, particularly those found in the ArchiMate Community[2] and The ArchiMate Cookbook[3].

On the *active structure*, previous research encompasses employees, customers, and competitors as business actors, as well as the business ecosystem as a business collaboration on the business layer. Software systems were modeled as application components on the application layer. On the physical layer, the firm location is modeled as a facility.

In the *behavior aspect*, the strategy layer comprises competencies of the employees as capabilities [22], as well as marketing tactics and circular economy actions as courses of action. On the business layer, human resources practices and product development are depicted as business functions, as well as service requests as business interaction. In the application layer, there are two application interactions: feature requests and app issue reporting. On the implementation and migration layer are business model innovations, which are traces of architectural change over time.

In the *passive structure*, two resources can be modeled on the strategy layer, namely patents and brand personality. Further on the business layer, the product features of the enterprise and its competitive advantages can be modeled as products.

Most elements can be assigned to the *motivation layer*. Here, customers can be modeled as stakeholders, driven by their satisfaction, assessable through continuance intention and recalls. A similar logic can be applied to (retail) investors, driven by their sentiment and assessable via funding data. Corporate values can be modeled as principles that shape the organizational culture, serving as a driving force for the enterprise. Simple goal-outcome relationships can be observed in the context of supply chain risks and downtime during crises, as well as in sustainability and corporate environmental performance.

It can be seen that a considerable amount of EA artifacts can be modeled from web sources. Notably, the motivation layer is sufficiently covered by existing research. There is a research gap in gaining insights into the technology layer from web sources. I argue that this is likely because of study designs that favor internal data. Here, job advertisements, especially those for IT-related roles, can provide insights into the applications and technologies used by an enterprise. Additionally, the potential of

[2]https://community.opengroup.org/archimate-community
[3]https://www.hosiaisluoma.fi/blog/archimate

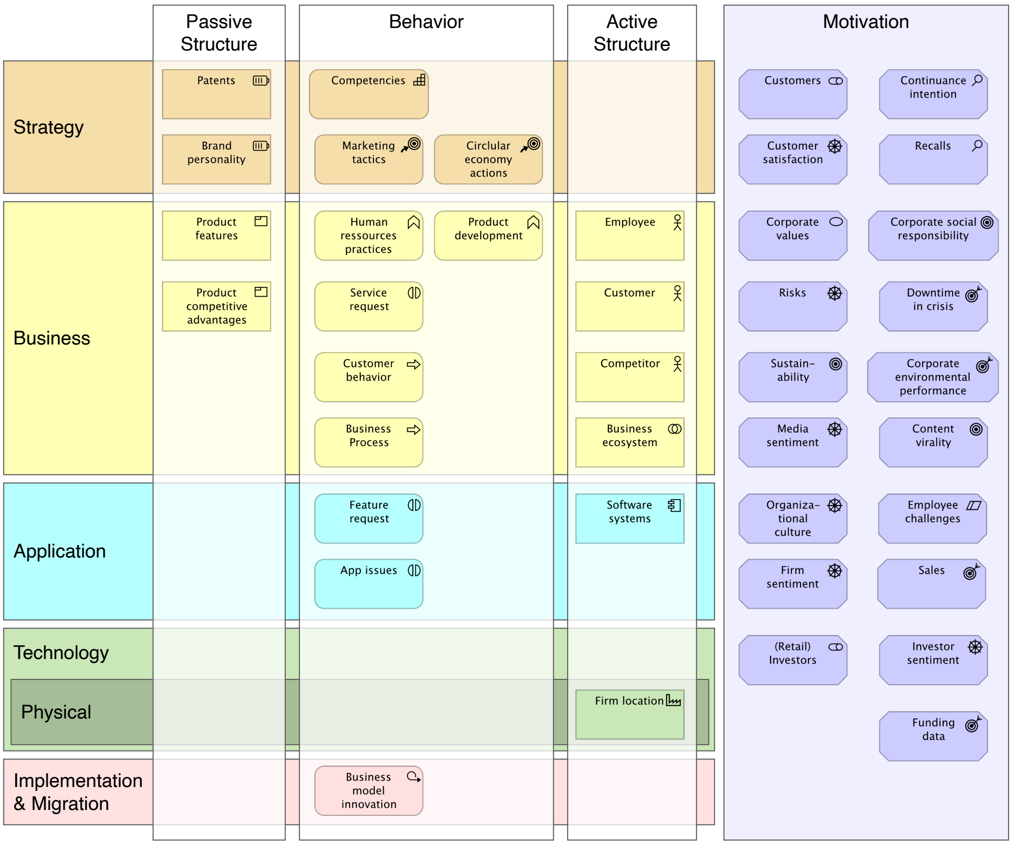


**Figure 2:** Enterprise-related insights mapped to EA artifacts.

corporate social media (e.g., LinkedIn, Glassdoor) is not yet fully utilized. Here, business actors, roles, and processes could be derived.

### 2.3.3. Ontological Perspective

The review found that, to date, there is no method for integrating enterprise-related web data into EA. However, some studies have applied ontologies to understand the nature of web data better. As illustrated in table 2, the majority of prominent social media platforms have been described ontologically. No study utilized a foundational ontology, which would have enabled mapping to other domains, particularly the EA domain. Given the abundance of ontology-based approaches, I will elaborate further on this thinking in the subsequent chapter.

**Table 2**
Web ontologies.

| Web source | Enterprise-related web data | Records |
|---|---|---|
| Crunchbase | Firm profiles | [61, 63] |
| Glassdoor | Employee reviews | [59] |
| Google Play Store, AppFigures | App reviews | [64, 65] |
| LinkedIn | User profiles | [61][23] |
| P2PEye (P2P information provider) | User comments | [66] |
| Soyoung (Healthcare platform) | Firm-generated content | [67] |
| TripAdvisor | Guest reviews | [68] |
| Twitter | Firm-generated content | [69] |
| Wayback Machine | Firm websites | [28] |
| Zomato.com (Food delivery website) | User reviews | [70] |
| autohome.com, Bitauto, Toyota Nation (Automotive forums) | User reviews | [51, 71] |

**Table 3**
Insights by originator, content type, and source.

| Originator (Who) | Content type (How) | Source (Where) | Enterprise-related insight (What) |
|---|---|---|---|
| (Retail) Investor | Social media post | Bloomberg [72], StockTwits [41], Twitter [41, 73], Reddit [73] | Investor sentiment [72, 41, 73] |
| Customer | Social media post | Twitter [74, 75, 52, 76, 77, 78, 79, 80], Facebook [81, 82, 83], Instagram [82], Reddit [84, 85, 86, 87], Box Office Mojo [88] | Customer satisfaction [52, 76, 88, 82, 83, 84, 89, 79, 86, 80, 87], Service requests [77], Product competitive advantages [81, 85, 75], Requirements [78]. |
| Customer | Review | TripAdvisor [90, 91, 68, 92, 93, 94], ordme.ir [95], airlinequality.com [96], Zomato.com [70], dianping.com [97], careopinion.org.uk [98], Company website [99], Yelp [100], bestbuy [101], P2PEye [66], Not mentioned [102] | Customer satisfaction [94, 90, 96, 99, 95, 68, 66, 102], Service requests [70, 91], Customer behavior [101, 92, 93] |
| Customer | Forum post | Bitauto [71, 51], autohome.com [51], Slashdot [50], Schneier on Security [50], Not mentioned [103] | Customers [71], Product competitive advantages [50, 51], Customer satisfaction [103] |
| Customer | App review | Apple App Store [104, 105, 106, 107, 108], Google Play Store [109, 53, 110, 105, 111, 112, 113, 114, 106, 107, 115, 108, 116, 117], SwiftKey [64], AppFigures [64], Github [118], MapMyRun [119] | Feature requests [105, 111, 112, 113, 65, 114, 104, 53], App issues [64, 109], Continuance intention [106], Customer satisfaction [107, 115, 108, 116, 117], Customer behavior [110, 119] |
| Employee | Social media post | Reddit [62], Twitter [46, 120], Glassdoor [120], LinkedIn [61] | Employee challenges [62], Business ecosystem [46], Brand personality [120], Competences [61] |
| Employee | Review | Glassdoor [59] | Organizational culture [59] |
| Firm | Social media post | Facebook [56, 42, 57, 121], LinkedIn [122, 121], Twitter [39, 56, 123, 124, 43, 58, 125, 121, 126, 55], Instagram [122, 126], VK.com [127], WeChat [56], Weibo [56], YouTube [123], Douyin [128] | Content virality [69, 56], Marketing tactics [125, 121, 122, 126, 39, 127, 123, 124], Brand personality [55], Downtime in crisis [42], Corporate values [128], Circular economy actions [43], Business model innovations [58, 57] |
| Firm | Press release | SEC Edgar [129], Factiva (Dow Jones) [37], eqs-news.com [130] | Product competitive advantages [129], Marketing tactics [37], Firm sentiment [130] |
| Firm | Patent | European Patent Office [131] | Software systems [131] |
| Firm | Job advertisement | seek.com.au [60], Indeed [132, 133, 30, 134], monster.com [135], LinkedIn [136, 30], JournalismJobs.com [137], Profession.hu [30], catho.com.br [138] infojobs.com.br [138], hays.com [138], pagepersonnel.com.br [138], michaelpage.com.br [138], LA Joblist [29], ISJobs.com [29], Buslib-L [29], Jobline.hu [30], Careerjet.hu [30], University portal [139] | Human resources practices [60], Competencies (Rest) |
| Firm | Company website | Wayback Machine [28], Company website [48, 49, 54, 31, 140, 44, 32] | Sustainability [28], Product features [48, 49], Brand personality [54], Employees [31], Customer satisfaction [31, 140], Corporate social responsibility [44], Business processes [32], Sales [31] |
| Firm | Release Notes | Apple App Store [141] | Product development [141] |
| Government | Registry | NHTSA (government database) [40], US Toxic Releases inventory [142], German Commercial Registry [35], Denmark statistics centre [45] | Recalls [40], Corporate environmental performance [142], Business ecosystem [35, 45] |
| Government | National survey | Mexican Social Security Institute [34] | Customer satisfaction [34] |
| Platform | Product page | Wish.com [33] | Sales [33] |
| Press | News article | Asian Times [47], BBC [47], Thomson Reuters [47], Economic Times [36, 143], Financial Express [36], investing.com [38], GDELT Project [144] | Risks [36, 47], Media sentiment [38, 143], Product features [144] |
| Data aggregators | Firm profile | Crunchbase [61, 63, 145], Soyoung (healthcare platform) [67] | Funding data [63, 145], Competences [61], Customer satisfaction [67], Product competitive advantages [67] Firm location [67] |

## 3. Method for EA Traces Collection

Based on the results, I propose that EA traces consist of three components: Enterprise-related web data, EA artifacts, and an ontologically grounded mapping function. The three-step method for EA traces collection in figure 3 is proposed to map enterprise-related web data to EA artifacts. This aligns with the formal resource-based technique, as described in [24, p. 77]:

1. Each domain comprises numerous, occasionally competing ontologies. These ontologies must be integrated within a unified *domain ontology* to facilitate interconnections.
2. Given that ontologies often differ between domains, it is imperative to ensure that there is a common understanding of the relevant terms to facilitate semantic interoperability. This objective is facilitated by upper-level or *foundational ontologies*, such as the Unified Foundational Ontology (UFO) [25]. Therefore, the second step involves aligning each domain ontology with the same foundational ontology.
3. As a last step, entities and relationships from both the EA and web domain can be connected via a *mapping function.* This function, besides a technical linkage, employs ontological reasoning and interpretation. I argue that EA traces can be considered as 'big data' as described by [26]; therefore, his ten big data characteristics (e.g., incomplete, nonrepresentative, drifting, algorithmically confounded, dirty, sensitive) must be accounted for in the mapping function.

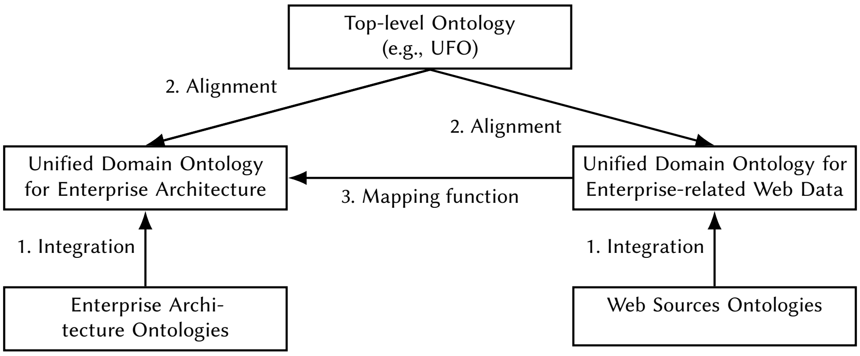


**Figure 3:** Method for EA traces collection.

## 4. Demonstration

The objective of this demonstration is to illustrate how ontological reasoning can facilitate the collection of EA traces. I drew upon two pre-existing domain ontologies from the review.

First, [22] developed an EA ontology to model employees' competencies in ArchiMate, utilizing UFO (Figure 4). UFO-A is employed to model endurants, defined as individual entities that exist in time. The individual in question is depicted as a substantial physical agent, operating independently of other constructs. In contrast, moments are predicated on the existence of different entities, i.e., the person. Consequently, human capabilities, comprising skills and personal competencies, are conceptualized in light of this understanding. These human capabilities subsequently manifest in the perdurant "task," which unfolds over time (UFO-B). The task itself is embedded in a capability context, which is modeled as a situation.

Second, LinkedIn emerged as a dominant source in the review, with an ontological representation of employee profiles already established [23]. Unfortunately, the conceptual model is not constructed with the guidance of a foundational ontology (such as UFO). Therefore, I needed to translate the conceptual model into UFO constructs (Figure 5). A LinkedIn profile typically consists of the person (employee) in question, their position within the organization, and their education, including relevant skills. Ontological reasoning makes us question the way we perceive web data. In the ontology proposed

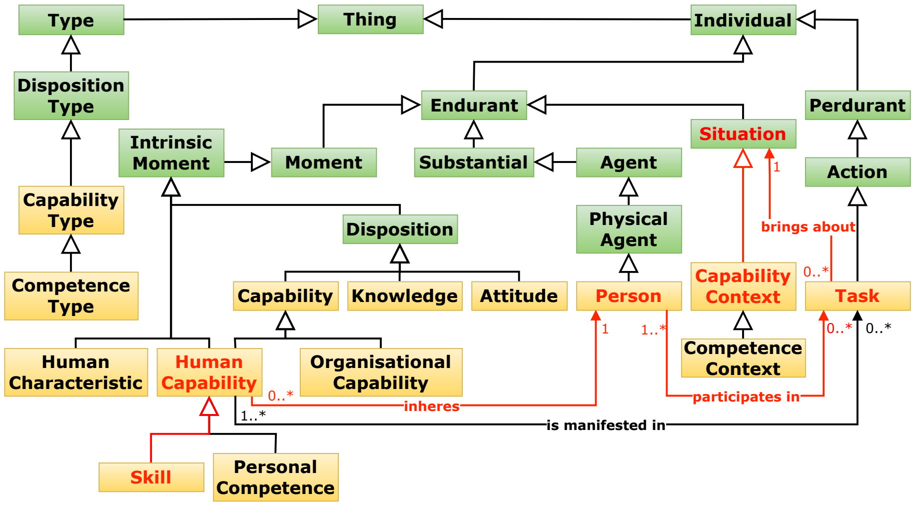


**Figure 4:** Modeling competency artifacts (yellow) in the EA domain ontology grounded on UFO concepts (green) [22].

by [23] the position can be understood either as a contractual entity ("Head of Marketing"), which would be modeled as a social relator in UFO, or as an temporal unfolding activity ("Leading a marketing strategy"), which would be modeled as an action. LinkedIn profiles typically contain both concepts, the position as a title and a description of the job. The latter fits better with the EA ontology and was therefore chosen. This does not imply that a representation as a social relator is not suitable for mapping EA artifacts. Instead, it highlights that the corresponding element is not present in the EA ontology by [22], and other EA ontologies should be integrated. For the organization, a similar ontological question arises. One would typically model the organization as a UFO institutional agent, as it comprises an enduring organizational identity, even when its EA changes. But in the context of competency modeling, [22] proposes the "competence context" to be a UFO situation, i.e., a snapshot of how EA holds together at a moment in time. These implications are important, as organizations must be modeled based on their configuration at a given moment in time.

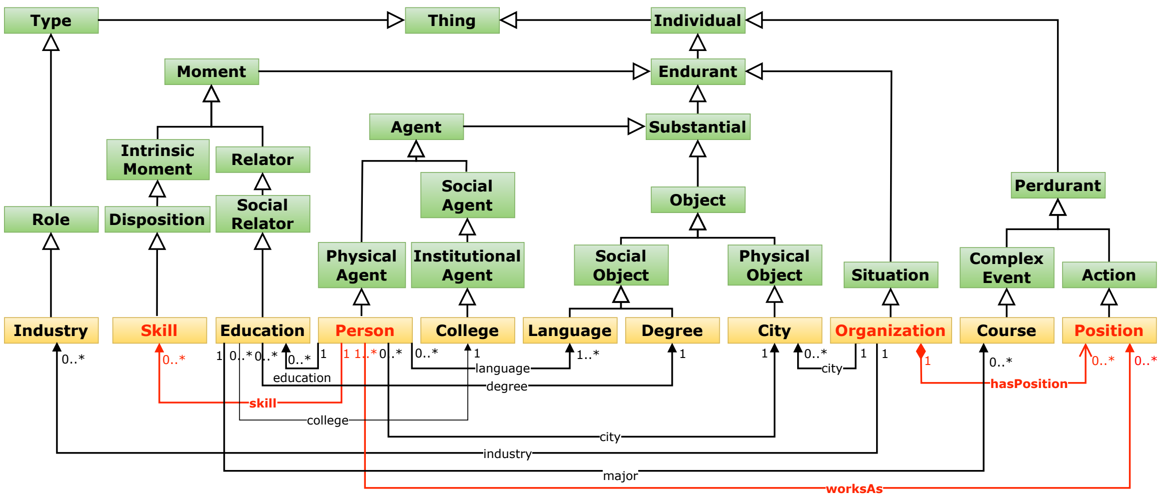


**Figure 5:** Modeling LinkedIn profiles in the web domain ontology (conceptual model by [23] [yellow] with custom UFO alignment [green]).

Given the alignment of these two domain ontologies with UFO, a mapping can be established, as illustrated in table 4. A mapping can be established between the person and their skills based on their disposition and physical agent, respectively. The position within the LinkedIn ontology can be mapped to a task in the EA ontology via an action, and the organization can be mapped to a capability context via a situation. Once the entities have been identified, their relationships can be mapped accordingly. The relationships concerning the skill inheritance are modeled in reverse. According to the LinkedIn ontology, a person possesses a skill, whereas, following the EA ontology, a skill is inherent to a person. In the LinkedIn ontology, an organization is defined as having a position. Conversely, in the EA ontology, a task is defined as bringing about a capability context. The relationship between a person and their position/task is expressed similarly in both cases, with the difference being in the choice of wording: "worksAs" for the LinkedIn and "participates in" for the EA ontology.

**Table 4**
Ontology mapping between EA and LinkedIn domain ontologies.

| LinkedIn Ontology concept | EA Ontology concept | UFO upper class | ArchiMate construct |
|---|---|---|---|
| **Entities** | | | |
| Skill | Skill | Disposition | Capability (Strategy layer) |
| Person | Person | Physical Agent | Business actor (Business layer) |
| Position | Task | Action | Business process (Business layer) |
| Organization | Capability Context | Situation | Plateau (Implementation & Migration) |
| **Relations** | | | |
| skill | inheres (reverses) | - | Association |
| hasPosition | brings about (reverse) | - | Composition |
| worksAs | participates in | - | Assignment |

Once the domain ontologies have been mapped, the EA trace is established and can be visualized in an EA framework. The translation proposed by [22] was employed to model the competency artifacts into ArchiMate constructs, as illustrated in figure 6. At the core of this EA trace lies the person, conceptualized as a business actor within the business layer. The person possesses skills, conceptualized as capabilities on the strategy layer, and engages in a task, modeled as an action on the business layer. The task results in the establishment of a capability context that is represented as a plateau on the implementation and migration layer. This demonstration serves as a critical link between the EA traces concept, the proposed method for collection, and the research agenda in the subsequent chapter.

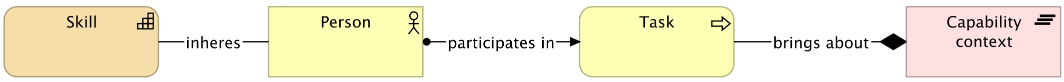


**Figure 6:** EA trace of LinkedIn profiles.

## 5. Conclusion

To date, only [32] and [17] have explicitly modeled EA from web sources. While those approaches focused solely on mapping EA artifacts from company websites, this paper introduces the EA traces concept for theoretical grounding and a method for their collection. Thereby, the paper builds upon ontology-based artifact retrieval approaches with internal sources [27], aiming for a conceptualization to integrate *publicly* available enterprise-related web data for faster modeling practices.

I posit that ontology thinking, particularly when guided by a robust foundational ontology such as UFO, can play a pivotal role in the development of EA models that are well aligned with web-sourced data. It is important to acknowledge that there is room for interpretation when integrating enterprise-related web data and EA artifacts. A three-step method is therefore designed to define the scope for interpretation. The approach was demonstrated by sourcing competency artifacts from LinkedIn profiles, thereby showing its feasibility.

The EA traces concept comes with some apparent challenges that future research needs to address:

1. Further development of an EA traces collection system is needed to capture a complete enterprise. Thereby, the mapping function needs to be extended to account not only for ontological reasoning but also to include countermeasures for deliberate manipulation of web content. The mapping function should be validated with the firms in question and be integrated with internal (private) data to expand the EA systems currently in use.
2. Data accessibility (APIs, scraping feasibility) skews research coverage toward platforms (e.g., Twitter, App Stores). I therefore propose other web sources to be investigated with an ontological perspective in mind.
3. Ethical and privacy concerns related to collecting employee or customer-generated content need to be addressed. Although this content is deliberately put online by users, responsible handling is a matter of course to avoid creating an 'the EA never forgets' scenario.
4. Once an EA traces collection system is built, further applications of EA traces can be investigated. This paper examined the concept of EA traces through a part of the 5W1H format (Who, How, Where, What), demonstrating a solid foundation for the method proposed. Looking ahead, EA traces can be expanded to incorporate temporal (When) and rationale (Why) facets. Integrating these could further enable meaningful interpretation of the architecture obtained via EA traces, ultimately paving the way for advanced EA mining systems. Since I propose EA can be (to a degree) generated from public sources, this implies that companies cannot only build EA for their own enterprise but also for competitors. This hints towards a new systematic approach for building large competitive intelligence systems.

In summary, the concept of EA traces introduced in this paper proposes a new perspective in the enterprise modeling domain, sparking potential for future research.

## Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

## References


[1] H. Holm, M. Buschle, R. Lagerström, M. Ekstedt, Automatic data collection for enterprise architecture models, Software & Systems Modeling 13 (2012) 825–841. doi:10.1007/s10270-012-0252-1.

[2] M. Farwick, C. M. Schweda, R. Breu, I. Hanschke, A situational method for semi-automated enterprise architecture documentation, Software & Systems Modeling 15 (2014) 397–426. doi:10.1007/s10270-014-0407-3.

[3] P. Johnson, M. Ekstedt, R. Lagerstrom, Automatic probabilistic enterprise it architecture modeling: A dynamic bayesian networks approach, in: 2016 IEEE 20th International Enterprise Distributed Object Computing Workshop (EDOCW), IEEE, 2016, pp. 1–8. doi:10.1109/edocw.2016.7584351.

[4] M. Osenberg, M. Langermeier, B. Bauer, Using Semantic Web Technologies for Enterprise Architecture Analysis, Springer International Publishing, Cham, 2015, pp. 668–682. doi:10.1007/978-3-319-18818-8_41.

[5] J. Dahlke, S. Schmidt, D. Lenz, J. Kinne, R. Dehghan, M. Abbasiharofteh, M. Schütz, L. Kriesch, H. Hottenrott, U. N. Kanilmaz, N. Grashof, A. Hajikhani, L. Liu, M. Riccaboni, P.-A. Balland, M. Wörter, C. Rammer, The WebAI paradigm of innovation research: Extracting insight from organizational web data through AI, ZEW Discussion Papers 25-019, ZEW - Leibniz-Zentrum für Europäische Wirtschaftsforschung, Mannheim, 2025. URL: https://hdl.handle.net/10419/319890.

[6] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, M. Van Alstyne, Computational social science, Science 323 (2009) 721–723. doi:10.1126/science.1167742.

[7] J. Ohme, T. Araujo, L. Boeschoten, D. Freelon, N. Ram, B. B. Reeves, T. N. Robinson, Digital Trace Data Collection for Social Media Effects Research: APIs, Data Donation, and (Screen) Tracking, Communication Methods and Measures 18 (2023) 124–141. doi:10.1080/19312458.2023.2181319.
[8] R. J. Torraco, Writing integrative literature reviews: Guidelines and examples, Human Resource Development Review 4 (2005) 356–367. doi:10.1177/1534484305278283.
[9] X. Xu, Studying social networks in the age of computational social science, EPJ Data Science 12 (2023). doi:10.1140/epjds/s13688-023-00436-z.
[10] S. A. Golder, M. W. Macy, Digital footprints: Opportunities and challenges for online social research, Annual Review of Sociology 40 (2014) 129–152. doi:10.1146/annurev-soc-071913-043145.
[11] M. Luca, User-Generated Content and Social Media, Handbook of Media Economics, Elsevier, 2015, pp. 563–592. doi:10.1016/b978-0-444-63685-0.00012-7.
[12] J. Howison, A. Wiggins, K. Crowston, Validity issues in the use of social network analysis with digital trace data, Journal of the Association for Information Systems 12 (2011) 767–797. doi:10.17705/1jais.00282.
[13] T. Grisold, W. Kremser, J. Mendling, J. Recker, J. vom Brocke, B. Wurm, Generating impactful situated explanations through digital trace data, Journal of Information Technology 39 (2024) 2–18. doi:10.1177/02683962231208724.
[14] S. Karanasios, D. Thakker, L. Lau, D. Allen, V. Dimitrova, A. Norman, Making sense of digital traces: An activity theory driven ontological approach, Journal of the American Society for Information Science and Technology 64 (2013) 2452–2467. doi:10.1002/asi.22935.
[15] D. Freelon, On the interpretation of digital trace data in communication and social computing research, Journal of Broadcasting & Electronic Media 58 (2014) 59–75. doi:10.1080/08838151.2013.875018.
[16] R. Boulland, T. Bourveau, M. Breuer, Corporate websites: A new measure of voluntary disclosure, SSRN Electronic Journal (2019). doi:10.2139/ssrn.3816623.
[17] A.-M. Ghiran, R. A. Buchmann, C.-C. Osman, D. Karagiannis, Streamlining Structured Data Markup and Agile Modelling Methods, Lecture Notes in Business Information Processing, Springer International Publishing, Cham, 2017, pp. 331–340. doi:10.1007/978-3-319-70241-4_22.
[18] M. A. Cronin, E. George, The why and how of the integrative review, Organizational Research Methods 26 (2020) 168–192. doi:10.1177/1094428120935507.
[19] H. M. Cooper, Organizing knowledge syntheses: A taxonomy of literature reviews, Knowledge in Society 1 (1988) 104–126. doi:10.1007/bf03177550.
[20] A. Carrera-Rivera, W. Ochoa, F. Larrinaga, G. Lasa, How-to conduct a systematic literature review: A quick guide for computer science research, MethodsX 9 (2022) 101895. doi:10.1016/j.mex.2022.101895.
[21] R. Winter, R. Fischer, Essential layers, artifacts, and dependencies of enterprise architecture, in: 2006 10th IEEE International Enterprise Distributed Object Computing Conference Workshops (EDOCW'06), IEEE, 2006, pp. 30–30. doi:10.1109/edocw.2006.33.
[22] R. F. Calhau, J. P. A. Almeida, S. Kokkula, G. Guizzardi, Modeling competences in enterprise architecture: from knowledge, skills, and attitudes to organizational capabilities, Software and Systems Modeling 23 (2024) 559–598. doi:10.1007/s10270-024-01151-7.
[23] J. Li, V. Wade, M. Sah, Developing knowledge models of social media: A case study on linkedin, Open Journal of Semantic Web 1 (2014) 1–24.
[24] J. Euzenat, P. Shvaiko, Ontology matching, 2nd ed., Springer-Verlag, Heidelberg (DE), 2013.
[25] G. Guizzardi, A. Botti Benevides, C. M. Fonseca, D. Porello, J. P. A. Almeida, T. Prince Sales, Ufo: Unified foundational ontology, Applied Ontology 17 (2022) 167–210. doi:10.3233/ao-210256.
[26] M. J. Salganik, Bit by Bit: Social Research in the Digital Age, Princeton University Press, Princeton, NJ, USA, 2019.
[27] C. R. Pinheiro, S. L. P. D. Guerreiro, H. S. Mamede, A lightweight ontology for enterprise architecture mining of api gateway logs, IEEE Access 12 (2024) 128585–128601. doi:10.1109/ACCESS.2024.3456119.