

Extending Crowdsourcing Provenance Models to Represent AI-Assisted Contributions

Samaneh Kiamanesh¹

¹Memorial University of Newfoundland, St. John's, NL A1C 5S7, Canada

Abstract

Crowdsourcing datasets usually identify who submitted an item, but not how it was produced. This matters when a worker uses generative artificial intelligence (GenAI) to draft, revise, or assess a contribution. Existing studies document both self-initiated and designed GenAI use but record it in different ways. General provenance standards represent agents, activities, and artifacts without providing shared crowdsourcing terms for assistance roles, disclosure, evidence, or version-specific verification. This study proposes a lightweight specialization of the World Wide Web Consortium's PROV Ontology (PROV-O) for AI-assisted crowd contributions. Following a design science research approach, a targeted evidence review, three scenarios, competency questions, and representational analysis will guide its development and comparison with base PROV-O. An initial review of 25 sources has identified eleven candidate constructs and nine competency questions. Rather than detecting AI use or assigning a universal trust score, the model aims to provide a transparent record for later task-specific quality judgments.

Keywords

Crowdsourcing, Generative Artificial Intelligence, provenance, PROV-O, conceptual modeling, verification

1. Introduction

Crowdsourcing distributes tasks to a group of contributors, usually through an online platform, to collect judgments, labels, observations, or created content on a scale. Because its value depends on whether the contributions fit the requester's purpose, quality research has examined tasks, workers, answers, workflows, and corrective actions [1, 2]. This work often assumes that the identified worker is also the main producer of the item.

Generative artificial intelligence (GenAI) makes that assumption less reliable. A worker may use a large language model for ideation, drafting, rewriting, translation, label recommendation, or critique. The worker may disclose this use, the platform may observe it, or the requester may infer it later. Yet the dataset may retain only one answer attributed to one person, hiding a human-AI process that can affect compliance, accountability, reproducibility, and quality assessment.

For example, two workers may submit similar explanations in a misinformation-assessment task. One writes independently; the other asks a large language model to rewrite a draft, edits the suggestion, and submits it. A dataset storing only the final text and worker identifier treats them alike. A provenance record can instead identify the AI activity, intermediate and final versions, available interaction evidence, and the version later verified.

This is already happening in crowdwork. Veselovsky et al. [3] estimated that 33–46% of usable submissions in one summarization task involved large language models. Christoforou et al. [4] also found self-initiated use across three platforms, with substantial platform differences. In designed workflows, AI labels and explanations changed workers' judgments and search behavior [5], while Yeh et al. [6] recorded worker drafts, AI suggestions, human revisions, and expert assessments. These findings show why an "AI used: yes/no" field is not enough.

Provenance is a useful starting point because it records the entities, activities, and agents involved in producing data. It has already supported reproducibility and quality analysis in crowdsourcing [7], and the World Wide Web Consortium's PROV Ontology (PROV-O) provides interoperable concepts

ER2026: Companion Proceedings of the 45th International Conference on Conceptual Modeling: Doctoral Consortium, Tutorials, Posters and Demos, October 5–8, 2026, St. John's, NL, Canada.



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

for people, software agents, activities, entities, associations, and roles [8]. However, it does not define crowdsourcing-specific assistance, disclosure evidence, or version-specific verification. Similar situations may therefore be encoded in valid but inconsistent ways.

This study develops a small, crowdsourcing-specific PROV-O profile for individual contributions, particularly in open-ended content and judgment tasks. It connects AI assistance, assertions about that assistance, supporting evidence, and verification of a particular contribution version. Observable activities are recorded first; requester policy determines later labels such as human-only, AI-assisted, or AI-generated.

The remainder of the paper defines the problem and scope (Section 2), reviews related work (Section 3), and states the questions and objectives (Section 4). Sections 5-6 present the model and method; Sections 7-9 report progress, contributions, and timeline; and Sections 10-11 discuss limitations, open questions, and conclusions.

2. Research Problem and Scope

The issue is not simply whether AI was used. A useful record should explain what the AI did, when it entered the process, what evidence supports that account, and which version was checked. Prompts without defined roles are insufficient, as is an “AI-assisted” label separated from its activity and evidence. Incompatible local schemas make it difficult to compare datasets, trace responsibility, or identify the version that was verified.

Three distinctions matter. Provenance provides evidence but does not prove quality. Disclosure is not detection because a worker statement, interface log, and detector inference have different status. Finally, assistance type is often more informative than a universal authorship label: drafting, translation, critique, and recommendation create different dependencies. Roles, activities, versions, and evidence should therefore be recorded before a requester classifies the result.

The initial study does not attempt to detect undisclosed AI use, calculate a universal quality score, redesign a platform, or model every form of human–AI collaboration.

The profile succeeds if it represents the selected scenarios consistently and answers its defined questions; it need not decide whether a contribution should be accepted.

This study is a foundational phase of a broader doctoral program on credibility-aware adaptive verification for crowdsourced data quality, not the full doctoral project. The profile establishes the provenance layer needed to distinguish human and AI activities and connect assertions, evidence, versions, and verification outcomes. Later studies will examine how these records can support contributor-credibility assessment, prioritize task-specific verification, and inform data-quality decisions across platforms and task types.

3. Related Work

3.1. Generative Artificial Intelligence in Crowdwork

GenAI crowdwork studies take two broad approaches. The first observes self-initiated use. Veselovsky et al. [3] inferred model use from text and interaction traces, while Christoforou et al. [4] compared worker self-reports across platforms. Both highlight contribution origin and disclosure, but rely on study-specific, often binary measures rather than a reusable representation of AI activities, roles, artifacts, and evidence.

The second approach intentionally includes AI in the workflow. Xu et al. [5] showed that labels and explanations can affect confidence, effort, and reliance differently. Yeh et al. [6] captured worker drafts, AI suggestions, human revisions, final comments, and expert assessments. These cases show why assistance role and version history matter, although their traces remain tied to individual study designs.

3.2. Crowdsourcing Provenance and Quality Control

The closest crowdsourcing-provenance predecessor is Martinez-Ortiz et al. [7], who mapped CrowdTruth artifacts, transformations, agents, parameters, and platform branches to the Open Provenance Model. Their approach supports workflow comparison and quality analysis but predates GenAI and does not describe AI activity inside a contribution attributed to one worker.

PROV-O is a suitable backbone [8]: a contribution can be a prov:Entity, production a prov:Activity, a worker a prov:Person, and an AI system a prov:SoftwareAgent. Qualified associations can assign prov:Role values. However, the crowdsourcing lifecycle, assistance vocabulary, disclosure status, evidence requirements, and verification outcomes remain open to local interpretation. The proposal specializes rather than replaces PROV-O. The PROV notation (PROV-N) may later provide compact example traces [9].

Quality research explains why this trace matters. Daniel et al. [2] organize quality attributes, assessments, and assurance actions across tasks, workers, outputs, requesters, and platforms. Chai et al. [1] emphasize the interaction among task, worker, answer, and workflow context. The proposed profile complements these frameworks by linking an assessment and outcome to the exact contribution version and production evidence concerned.

3.3. Conceptual Modeling of Crowdsourcing and Human–AI Collaboration

Conceptual modeling research suggests that crowd data should accommodate varied contributor perspectives without unnecessary complexity. Lightweight instance-oriented modeling [10] and guidelines for crowd knowledge [11] support a contribution-centered, parsimonious profile. Related studies show that modeling choices can affect completeness and argue for broadly usable principles [12, 13].

The Contextual Collaboration AI Ontology (CCAI) is the closest adjacent model identified so far [14]. It represents human and AI agents, tasks, resources, constraints, prompts, outputs, responsibilities, and roles while reusing PROV-O. The present study therefore does not claim to be the first human–AI provenance model. Its narrower contribution concerns individual crowd contributions, requester policy, disclosure or inference, versions, and requester-facing verification. Suitable CCAI terms will be reused or aligned rather than duplicated.

The evaluation combines competency questions with representational analysis. Competency questions define scope and test what users can ask of an ontology [15]. Representational analysis checks for missing concepts and overloaded constructs [16, 17]. Together they provide a feasible way to expose weaknesses in an early-stage model.

3.4. Research Gap

The gap is modest but useful. The 25-source corpus includes GenAI crowdwork studies, generic provenance mechanisms, and a broad human–AI ontology, but no lightweight crowdsourcing-specific PROV-O profile that combines AI assistance roles, the basis and evidence for assistance assertions, and verification of a particular contribution version. This is a corpus-based conclusion, not a universal claim that no related model exists. Targeted searching will continue during development.

4. Research Questions and Objectives

The study is guided by three research questions:

RQ1. What concepts and relationships are needed to represent whether and how GenAI assisted the production of an individual crowdsourced contribution?

RQ2. How can PROV-O be specialized to distinguish human-only, AI-assisted, and AI-generated contributions while recording assistance assertions, evidence, and verification?

RQ3. To what extent can the proposed specialization answer competency questions and represent common contribution and verification scenarios more completely and clearly than base PROV-O alone?

The objectives are to identify the required concepts, define a small PROV-O specialization, build representative traces, and compare the profile with base PROV-O. The study assumes no universal threshold for the three contribution labels: the profile records activities, roles, versions, assertions, and evidence, while requester policy determines classification.

5. Preliminary Conceptual Model

5.1. Design Principles

The profile reuses PROV-O where possible, records activities before labels, links derivation and verification to exact versions, distinguishes declarations from observations and inferences, and records verification without treating provenance as proof or as a universal trust score.

5.2. Candidate Constructs and Mappings

The working vocabulary organizes the proposed profile around task context, contribution production, evidence, and verification. A CrowdsourcingTask is anchored in prov:Plan, while a CrowdWorker and an AISystem are represented as prov:Person and prov:SoftwareAgent. The AISystem denotes the active service or runtime involved in the activity; model or version identifiers can be recorded separately when available. ContributionProduction, AIAssistanceActivity, and VerificationActivity specialize prov:Activity. ContributionVersion, AssistanceAssertion, EvidenceArtifact, and VerificationOutcome specialize prov:Entity, with versions connected through prov:wasRevisionOf. AssistanceAssertion records the asserted status of AI involvement—declared, observed, inferred, not declared, or unknown—and its source or basis. AssistanceRole, anchored in prov:Role through a qualified association, specifies ideation, drafting, rewriting, translation, critique, or recommendation. An artifact such as a prompt, output, or log may be both a production artifact and supporting evidence; these uses are contextual rather than mutually exclusive. A construct will be retained only if it supports a competency question or addresses a clear gap in base PROV-O.

5.3. Relationships and Interpretation

Figure 1 shows the main relationships. A task governs contribution production through a plan, while the worker remains associated with production and submission. If an AI system participates, a qualified association records its AssistanceRole in an AIAssistanceActivity. The activity uses or generates EvidenceArtifacts, and contribution versions are linked by revision or derivation. An AssistanceAssertion records whether involvement was declared, observed, inferred, not declared, or unknown, together with the source of that assertion. A VerificationActivity assesses a specific version while considering the assertion and supporting evidence, then generates a VerificationOutcome.

Missing AI evidence does not prove human-only production, and the presence of an AI agent does not automatically mean “AI-generated.” These labels should be derived from the recorded activities under requester policy, allowing one trace to support different task rules without a universal threshold.

5.4. Evaluation Scenarios

Three literature-grounded scenarios test the profile [4, 3, 5, 6]. S1 records production without reported AI assistance but avoids treating missing evidence as certainty. In S2, later evidence suggests undeclared AI use; the worker’s statement, inference, evidence, and submission responsibility remain distinct. S3 records a worker draft, AI suggestion, human revision, submission, and expert or hybrid verification. It tests roles, version history, partial adoption, and verification of the final output.

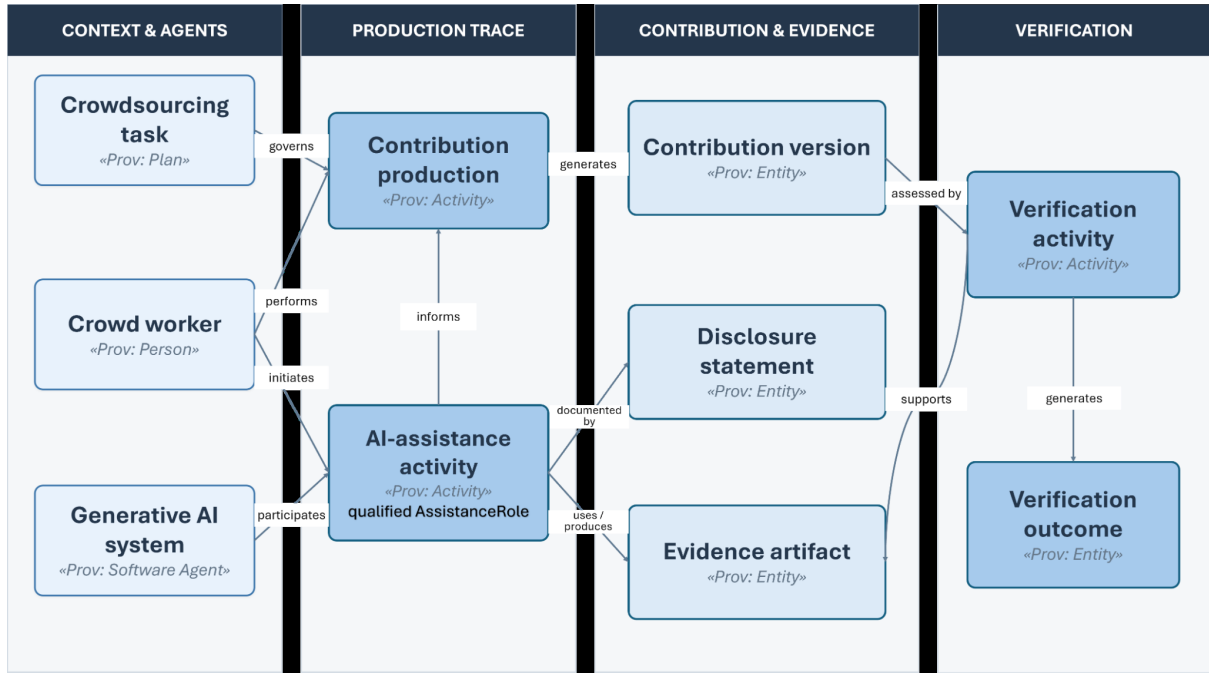


Figure 1: Simplified preliminary model. The profile is a specialization of PROV-O that adds crowdsourcing semantics for assistance assertions, evidence, and version-specific verification.

6. Design Science Research Method and Evaluation Plan

The study follows design science research, which develops and evaluates an artifact for a relevant information-systems problem [18]. The artifact is a lightweight, machine-readable PROV-O profile comprising classes, relationships, mappings, example traces, competency questions, and queries. Its context is open-ended crowdsourcing content and judgment tasks requiring transparent human-AI production records and version-specific verification. Phase 1 addresses problem identification and objectives, Phase 2 design, development, and demonstration, and Phase 3 evaluation and refinement; the paper and specification communicate the results.

6.1. Phase 1: Evidence-Informed Requirements

Phase 1 derives requirements from a matrix of 25 reviewed sources covering GenAI crowdwork, provenance, quality control, conceptual modeling, and model evaluation. The matrix records context, method, constructs, findings, limitations, and use in this project. It is a targeted evidence review for artifact development, not a claimed completed systematic review. Searches for human-AI provenance, disclosure schemas, and contribution models will continue, with an updated inclusion record.

Requirements become competency questions, constraints, and scenario facts, keeping each profile element traceable to a stated need.

6.2. Phase 2: PROV-O Specialization

Phase 2 defines the classes, relationships, assistance roles, assertion statuses, and minimum scenario traces. Every new term map to a PROV-O superclass, relation, or qualified association, with CCAI reuse or alignment considered where relevant [14]. The specification will include an ontology implementation, diagrams, examples, logical checks, and queries. General provenance tools should still interpret the core lineage.

6.3. Phase 3: Competency-Question and Representational Evaluation

Phase 3 represents each scenario with both base PROV-O and the proposed profile. The same competency questions are applied, with answers rated complete, partial, or not directly available. Complete means answerable from the representation without an undocumented local convention; partial means answerable only through an added convention or inference; and not directly available means that a required fact, construct, or relationship is absent. The comparison also records ambiguity and missing or overloaded constructs. The question is not whether PROV-O can encode the cases, but whether the profile makes them more consistent, explicit, and queryable.

A small formative expert review will assess clarity, scoped completeness, and likely usefulness. Experts in conceptual modeling, provenance, or crowdsourcing will inspect the specification and traces, identify ambiguous or unnecessary constructs, and assess whether the competency questions reflect practical information needs. This review refines the first artifact before later empirical studies; it does not establish platform-level effectiveness.

Table 1

Preliminary competency questions used to define scope and evaluate scenario representations.

ID	Competency question
CQ1	Who or what generated and submitted this contribution version?
CQ2	Was GenAI involved, and was its involvement declared, observed, inferred, not declared, or unknown?
CQ3	What role did the AI play: ideation, drafting, rewriting, translation, critique, or label recommendation?
CQ4	Which AI system and which prompt, input, output, or log artifacts were involved, when recorded?
CQ5	Which contribution version resulted from each human or AI activity, and how are the versions related?
CQ6	Which task instruction or policy governed whether assistance was permitted and what disclosure was required?
CQ7	What verification activity was performed, by whom or what, using which method and evidence?
CQ8	What outcome did verification produce, and to which contribution version and assistance trace does it apply?
CQ9	Can the path from task and source artifacts through human–AI production to submission and verification be reconstructed?

6.4. Evaluation Criteria

Evaluation considers competency-question coverage, explicit representation of role and evidence, avoidance of missing or unnecessary constructs, and parsimony without losing PROV-O interoperability.

7. Preliminary Results and Current Progress

The review has clarified the study’s direction. GenAI use may be hidden, self-reported, or designed [4, 3, 5, 6], and assistance form can affect process and output [5, 6]. PROV-O already supplies agents, activities, roles, and artifacts [8], while CCAI models human–AI context through provenance [14]. The contribution must therefore be a focused domain specialization rather than a new generic provenance foundation.

The current artifact contains eleven constructs, five principles, nine competency questions, and three scenarios. Its key choice is to record production activities before authorship labels, allowing different requester policies to interpret one trace. Treating assistance assertions and their evidence as provenance entities preserves the worker’s original statement when a later observation or inference is added.

Next, the profile and paired scenario data will be encoded, evaluated, and revised to remove constructs that add no clear value.

8. Expected Contributions

Expected contributions are a clearer distinction among AI participation, disclosure, evidence, and verification; a lightweight PROV-O profile with mappings and examples; a scenario-based evaluation; and a reusable basis for task-specific quality policies. The study also illustrates focused specialization of a general ontology and connects crowdsourcing provenance with quality records by treating AI assistance as accountable activities rather than a binary property of final text.

9. Research Plan and Timeline

The research plan separates the current ontology artifact from the later doctoral studies that will use its provenance records for credibility-aware verification and data-quality decisions.

Table 2

Research milestones, status, and planned schedule.

Stage	Primary Output	Status	Planned Period
M1	Evidence-to-requirement trace and finalized competency questions	In progress	Aug-Sep 2026
M2	PROV-O profile, mappings, queries, and example traces	Planned	Oct-Dec 2026
M3	Base PROV-O comparison and competency-question evaluation	Planned	Jan-Apr 2027
M4	Formative expert review and revised specification	Planned	May-Aug 2027
M5	Credibility-aware verification and empirical extension	Future phase	Sep 2027-Aug 2028

10. Anticipated Limitations, Risks, and Open Questions

The initial scenarios focus on text and judgment tasks, so later extensions may be needed for images, sensors, or autonomous agents. The profile can record the basis of a disclosure or inference but cannot establish ground truth when evidence is missing. To avoid duplicating PROV-O or CCAI, every new term must answer a competency question and have an explicit alignment. Analytical evaluation also cannot establish platform usability; expert feedback can provide early evidence, while implementation studies remain future work.

The symposium will seek feedback on three open questions: What is the smallest, useful extension that avoids duplicating PROV-O or CCAI? How should conflicting declarations, observations, and inferences be represented without implying ground truth? What evaluation beyond scenario and competency-question coverage would provide credible evidence of usefulness for later credibility-aware verification?

11. Conclusion

A worker identifier and final answer are incomplete when GenAI may have contributed. A useful record should show the human and AI activities, assistance role, assertion and its basis, supporting evidence, contribution versions, and verification. The proposed PROV-O profile brings these elements together and evaluates them through scenarios, competency questions, and representational analysis.

It makes human–AI production easier to inspect while leaving final quality judgments to the requester and task.

Declaration on Generative AI

During the preparation of this work, I used Grammarly for grammar and spelling checking. I independently reviewed, verified, and edited all content and took full responsibility for the publication's content.

References

- [1] L. Chai, H. Sun, J. Zhang, Quality control in open-ended crowdsourcing: A survey, *Frontiers of Computer Science* 20 (2026) Article 2006330.
- [2] F. Daniel, P. Kucherbaev, C. Cappiello, B. Benatallah, M. Allahbakhsh, Quality control in crowdsourcing: A survey of quality attributes, assessment techniques, and assurance actions, *ACM Computing Surveys* 51 (2018) Article 7, 1–40.
- [3] V. Veselovsky, M. H. Ribeiro, R. West, Artificial artificial artificial intelligence: Crowd workers widely use large language models for text production tasks, *arXiv*, 2023.
- [4] E. Christoforou, G. Demartini, J. Otterbacher, Generative ai in crowdwork for web and social media research: A survey of workers at three platforms, in: *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, 2024, pp. 2097–2103.
- [5] J. Xu, L. Han, S. Sadiq, G. Demartini, On the role of large language models in crowdsourcing misinformation assessment, in: *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, 2024, pp. 1674–1686. doi:10.1609/icwsm.v18i1.31417.
- [6] M. H. Yeh, R. Wan, T. H. Huang, CoCoLoFa: A dataset of news comments with common logical fallacies written by llm-assisted crowds, in: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2024, pp. 660–677.
- [7] C. Martinez-Ortiz, L. Aroyo, O. Inel, S. Champilomatis, A. Dumitrache, B. Timmermans, Provenance-driven representation of crowdsourcing data for efficient data analysis, in: *2015 IEEE 11th International Conference on e-Science*, IEEE, 2015, pp. 300–303.
- [8] T. Lebo, S. Sahoo, D. McGuinness, PROV-O: The PROV ontology (W3C recommendation), World Wide Web Consortium, 2013. URL: <https://www.w3.org/TR/prov-o/>.
- [9] L. Moreau, P. Missier, J. Cheney, S. Soiland-Reyes, PROV-N: The provenance notation (W3C recommendation), World Wide Web Consortium, 2013. URL: <https://www.w3.org/TR/prov-n/>.
- [10] R. Lukyanenko, J. Parsons, Lightweight conceptual modeling for crowdsourcing, in: W. Ng, V. C. Storey, J. C. Trujillo (Eds.), *Conceptual Modeling: ER 2013*, volume 8217 of *Lecture Notes in Computer Science*, Springer, 2013, pp. 508–511.
- [11] R. Lukyanenko, J. Parsons, Y. Wiersma, G. Wachinger, B. Huber, R. Meldt, Representing crowd knowledge: Guidelines for conceptual modeling of user-generated content, *Journal of the Association for Information Systems* 18 (2017) 297–339.
- [12] R. Lukyanenko, J. Parsons, Y. Wiersma, The impact of conceptual modeling on dataset completeness: A field experiment, in: *Proceedings of the Thirty Fifth International Conference on Information Systems (ICIS 2014)*, Association for Information Systems, 2014. URL: <https://aisel.aisnet.org/icis2014/proceedings/GeneralIS/29/>.
- [13] R. Lukyanenko, B. M. Samuel, J. Parsons, V. C. Storey, O. Pastor, A. Jabbari, Universal conceptual modeling: Principles, benefits, and an agenda for conceptual modeling research, *Software and Systems Modeling* 23 (2024) 1077–1100.
- [14] N. L. Lê, M. H. Abel, B. Laforge, From prompts to context: An ontology-driven framework for human–generative ai collaboration, *Information Systems Frontiers* (2026). Advance online publication.

- [15] G. K. Q. Monfardini, J. S. Salamon, M. P. Barcellos, Use of competency questions in ontology engineering: A survey, in: J. P. A. Almeida, J. Borbinha, G. Guizzardi, S. Link, J. Zdravkovic (Eds.), Conceptual Modeling: ER 2023, volume 14320 of *Lecture Notes in Computer Science*, Springer, 2023, pp. 45–64.
- [16] Y. Wand, R. Weber, On the ontological expressiveness of information systems analysis and design grammars, *Information Systems Journal* 3 (1993) 217–237.
- [17] Y. Wand, R. Weber, On the deep structure of information systems, *Information Systems Journal* 5 (1995) 203–223.
- [18] K. Peffers, T. Tuunanen, M. A. Rothenberger, S. Chatterjee, A design science research methodology for information systems research, *Journal of Management Information Systems* 24 (2007) 45–77.