

# Competitive Benchmarking: An IS Research Approach to Address Wicked Problems with Big Data and Analytics

Wolfgang Ketter<sup>\*</sup>, Markus Peters<sup>\*</sup>, John Collins<sup>\*\*</sup>, and Alok Gupta<sup>\*\*</sup>

<sup>\*</sup> Rotterdam School of Management, Erasmus University, Netherlands, [wketter@rsm.nl](mailto:wketter@rsm.nl), [peters@rsm.nl](mailto:peters@rsm.nl)

<sup>\*\*</sup> University of Minnesota, [jcollins@cs.umn.edu](mailto:jcollins@cs.umn.edu), [gupta037@umn.edu](mailto:gupta037@umn.edu)

**Abstract.** Wicked problems like sustainable energy and financial market stability are societal challenges that arise from complex socio-technical systems in which numerous social, economic, political, and technical factors interact. Understanding and mitigating them requires research methods that scale beyond the traditional areas of inquiry of Information Systems (IS) “individuals, organizations, and markets” and that deliver solutions in addition to insights. We describe an approach to address these challenges through Competitive Benchmarking (CB), a novel research method that helps interdisciplinary research communities to tackle complex challenges of societal scale by using different types of data from a variety of sources such as usage data from customers, production patterns from producers, public policy and regulatory constraints, etc. for a given instantiation. Further, the CB platform generates data that can be used to improve operational strategies and judge the effectiveness of regulatory regimes and policies. We describe our experience applying CB to the sustainable energy challenge in the Power Trading Agent Competition (Power TAC) in which more than a dozen research groups from around the world jointly devise, benchmark, and improve IS-based solutions.

**Keywords:** Benchmarking, Big Data Analytics, Design Science, Energy Information Systems, Research Competitions, Smart Grids, Sustainability, Virtual Worlds

## Introduction

“Wicked problems” ([Rittel and Webber, 1973](#)) like energy sustainability arise in complex socio-technical systems where numerous social, economic, political, and technical factors interact. The overall behavior of such a system cannot be explained by considering each of its parts in isolation, making it difficult to design targeted interventions that correct perceived misbehaviors of the system ([Kling, 2007](#)). Worse yet, even where promising interventions are known, the prohibitive cost of potential social negatives makes it impossible to thoroughly evaluate candidate interventions realistically and at scale.

We present a new conceptual and methodological approach by which IS research can begin to address such large scale, multifaceted, data-intensive problems. Our approach leverages data from a variety of

sources across multiple echelons<sup>1</sup> ([Lee and Whang, 1999](#)) and across multiple interconnected value chains, e.g., energy, housing, and transport to address problems such as the impact of retail competition and pricing policies in energy markets. We demonstrate that the collaborative combination of design science, behavioral research, and simulated competitive environments in our approach can be used to address large-scale, data-intensive, wicked problems. Such an interdisciplinary research approach requires efforts beyond the scope of a single methodological and/or modeling approach, and should be familiar ground for the Information Systems (IS) discipline with its rich tradition of studying and resolving socio-technical challenges for which solutions cannot be deduced from scientific principles alone ([Hevner and Chatterjee, 2010](#)). Events like electrical blackouts or recent financial market flash crashes have left the public wondering whether we may be becoming critically dependent on large-scale IT systems that we simply do not understand ([Cliff and Northrop, 2012](#)). But even though the IS discipline seems well-positioned to engage in these debates, its impact on large scale problem solving that address and resolve wicked problems has remained limited ([Lucas Jr et al., 2013](#); [Straub and Ang, 2011](#); [Schoder et al., 2014](#)). We make three contributions to this end:

First, we characterize the difficulties that wicked problems of societal scale pose to IS researchers. We contend that several obstacles limit the ability of current research methods to tackle problems of essential complexity that are large in scale and scope, that are currently unrealized, that progress at a rapid pace, and for which the social costs of erroneous interventions are prohibitive.

Second, we propose **Competitive Benchmarking** (CB) to address these obstacles. Our method draws on the authors' deep experience in the Trading Agents community, including designing and implementing the Power TAC scenario and organizing competitions. It emphasizes the importance of rich problem representations that are jointly developed among stakeholders and researchers, and it leads to actionable research results complete with comprehensive supporting data. Competitive Benchmarking supports analytical and behavioral IS research (insights) and design science research (solutions). We define CB as:

an approach to addressing a real-world wicked problem that is beyond the capacity of a single discipline or research team, by developing a shared paradigm consisting of problem definitions, vocabulary, and research questions; representing it in a tangible, open simulation platform; evaluating potential solutions from a wide range of researchers in direct competition with each other; and an ongoing process that continually updates the paradigm and platform to represent updated understanding of the real-world challenges and platform performance.

---

<sup>1</sup>The management science literature talks about “decentralized multi-echelon supply chains” ([Lee and Whang, 1999](#)) when referring to different, but connecting elements (and inventory) in the overall network.

Third, we apply Competitive Benchmarking in the **Power Trading Agent Competition** for research on sustainable energy systems (Power TAC, ([Ketter et al., 2013a](#), [2016](#))). Power TAC challenges researchers to design competing Energy Broker Agents ([Collins and Ketter, 2014](#)) as autonomous information systems that must translate continuous data streams into actionable results. Power TAC tests the notion that such entities can play a pivotal role as modern coordination mechanisms for sustainable energy systems specifically, and other smart market environments more generally ([Bichler et al., 2010](#)). To date, Power TAC has brought together more than a dozen research groups from various academic disciplines and stakeholders from utilities to customer lobby groups to design, evaluate, and improve both the Power TAC scenario and the competing Brokers.

We conclude with a detailed description of the process we used to build the Power TAC community, organize and conduct multiple competitions, and construct the associated IS artifacts and data repositories. This process follows an annual cycle of refinement based on feedback from participants and stakeholders, and on new and updated data resource.

## **Related Work**

CB borrows several tenets from the design science approach ([Hevner et al., 2004](#)). For example, we envision the design of each module to be based on a specific organizational or societal problem that is of interest to businesses, governments, or society in general. In contrast to Predictive Analytics ([Shmueli and Koppius, 2011](#)), which makes predictions about the future using data about the past and present, CB uses simulation driven by past and present data to test alternative future scenarios. The key difference lies in its emphasis on the interconnections between problems at the same and different echelons to mimic the real-world by leveraging ingenuity of a diversity of researchers. Repeated competitions then test the robustness of environment, for instance a market, its ruleset (policies), and models. Individual teams, entering to win the competition, try to, and often do, discover and exploit weaknesses or loopholes in policies and models to gain a superior position. This, in turn, allows designers to improve their designs and/or mitigate potential loopholes. The diversity of designs and the resulting exploits are difficult to achieve in traditional design science frameworks even when extensive simulations are used to test those designs. A much more rapid evolution of design occurs in a CB framework as compared to the traditional systems analysis and design approach.

Benchmarking has long been recognized as an important tool for improving products and organizational performance. Walter Chrysler regularly bought and disassembled new Oldsmobiles to

better understand his competition ([Shetty, 1993](#)), and Ford engineers allegedly anatomized some fifty German and Japanese cars before embarking on the construction of the popular Ford Taurus ([Mittelstaedt, 1992](#)). But, the key event that popularized benchmarking as a distinct concept among management practitioners and scholars was Xerox Corp.'s benchmarking-driven turnaround in the late seventies ([Garvin, 1993](#)). Today, a wide range of activities are recognized as benchmarking, ranging from informal comparisons within corporate boundaries to highly structured analyses of competitive postures across industries.

Competitive Benchmarking, as we define it here, is rooted in the competitive research approach pioneered by the Trading Agents community ([Greenwald and Stone, 2001](#); Collins et al., 2010a; [Wellman, 2011](#); [Ketter and Symeonidis, 2012](#)), which aims to deploy techniques from Artificial Intelligence and other computational disciplines to trading applications. Trading Agent Competitions (TAC) challenge researchers to devise software agents for complex, uncertain environments such as supply chains ([Arunachalam and Sadeh, 2005](#); Ketter et al., 2012) and advertisement auctions ([Jordan and Wellman, 2010](#)), to benchmark them in direct competitions with each other, and to improve them iteratively. This practice has been found to foster creativity, improve learning, and facilitate innovation based on deep introspection ([Garvin, 1993](#); [Shetty, 1993](#); [Drew, 1997](#)).

In simulation-based research, there is a tension between real-world fidelity and ease of statistical analysis. Though they have been inspired by interesting business problems, earlier TAC scenarios have focused on stability and abstraction to allow detailed statistical comparisons of agent behaviors across many competitions. CB (and Power TAC) instead focus on real-world relevance, and on a continuously evolving understanding of the challenge, and therefore may sacrifice the ability to compare the detailed performance of agents from one competition to the next. CB improves over TAC by providing human-system interaction facilities that can be used in training human decision-makers and in decision support studies. Such facilities are valuable in complex environments like financial markets, where training based on historical data streams “cannot readily model market impact ... [offering] essentially nothing toward understanding the current or future overall system-level dynamics ... [it] can tell you what happened, but not what might happen next, nor what might have happened instead” ([Cliff and Northrop, 2012](#)).

An interesting methodological question is how rigorous design theories can be derived from the comprehensive data generated by Power TAC competitions. Although a lively debate has been held on what constitutes a proper design theory, e.g., ([Walls et al., 1992](#); [Gregor and Jones, 2007](#); [Venable et al., 2012](#)), it is less clear how such a theory is best constructed starting from raw observational data. The Trading Agents community has addressed this issue using data-driven methods including descriptive

analyses ([Ketter et al., 2013c](#)), formal statistical or information-theoretical methods ([Andrews et al., 2009](#)), and empirical game theory ([Jordan et al., 2007](#)). We are currently evaluating their benefits for the derivation of principled IS design theories.

Scientific competitions such as those organized by Netflix ([Bell and Koren, 2007](#)) and Kaggle (<http://www.kaggle.com>) encourage participants to develop solutions for data mining, optimization, and forecasting problems ranging from movie preference correlations to disease spread analyses. They attract diverse communities of experts a variety of technical backgrounds, and produce data repositories that can be used to explore solution spaces and derive design theories. However, the artifacts developed for such scientific competitions do not interact directly with each other, and the one-shot nature of most such events precludes the collaborative analysis, learning, and iterative improvement process that is central to CB. Participants are limited to deploying promising techniques to prefabricated datasets provided by a self-interested sponsor, whereas identification and modeling issues remain out of their scope. Scientific competitions are therefore limited in their ability to produce insights and solutions for wicked problems for which the problem definition itself constitutes a significant hurdle ([Wagstaff, 2012](#)).

By contrast, work on Agent-based Computational Economics (ACE, [Tesfatsion, 2006b](#)) and Agent-based Virtual Worlds (ABVW, [Chaturvedi et al. 2011](#)) brings these modeling aspects to the foreground in an effort to evaluate possible futures of high-complexity environments, and potential paths to these futures, based on realistic assumptions. CB PLATFORMS are Virtual Worlds by definition, ideally constructed around relevant real-world data, and design guidelines like the involvement of citizen developers are important in their construction.<sup>24</sup> In contrast to ABVW, we make use of PLATFORMS as one of several components in an overarching method for IS research on wicked problems to alleviate the problem that:

[a]nalytical methods give elegant closed-form solutions to narrow, but well-defined, problems; empirical methods allow researchers to test theories at different levels of analyses; and computational methods allow researchers to build high fidelity simulations. However, none of these methods are particularly effective for studying large-scale problems ([Chaturvedi et al., 2011](#), p.682).

Beyond Virtual Worlds, CB adds the novel notion that software-based PLATFORMS can be used as the medium for capturing a community-created scientific paradigm, and as the infrastructure for a new type of competitive research process. The iterative, competitive nature of the PROCESS is essential in the context of wicked problems, because it brings the competitive co-evolution of artifacts into the laboratory, as well as the environmental complexity captured by regular ABVWs.

---

<sup>24</sup>More specifically, CB platforms are so-called Mirror Worlds, one of the two subtypes of ABVW ([Chaturvedi et al., 2011](#)).

Bringing elements of real-world evaluations into the laboratory is also prominent in the use of Serious Games for artifact evaluation ([Lang et al., 2009](#)) where participants engage in games that incorporate the artifact under study, e.g., a particular market mechanism. Similar to a CB PROCESS, these participants can evaluate the artifact more realistically than an isolated research group, since their diverse, creative behaviors will better pinpoint unintended design flaws. But Serious Games focus on human evaluations of a single artifact, whereas CB studies the competitive co-evolution of artifacts in complex environments. Moreover, unlike CB, Serious Games provide no tools for handling the scale and complexity inherent in research on wicked problems.

Table 1 summarizes the preceding discussion of related work.

### **Information Systems Research for Wicked Problems: Data and Design**

We set the scene for Competitive Benchmarking by first considering the difficulties that wicked problems pose to IS researchers. Two fundamental types of scientific inquiry can be distinguished in the IS discipline, both of which are important in resolving these challenges: behavioral research and design science research ([March and Smith, 1995](#); [Walls et al., 1992](#)). The research framework of [Hevner et al. \(2004\)](#) depicted in Figure 1 illustrates the interaction between the two. The circled numbers in the figure are referenced in the following text.

An IS research effort might start with the realization that IT can improve the effectiveness or efficiency of a particular socio-technical system, such as an organization's use of IT, or that of a whole society ①. If the goal of the research effort is to describe or explain phenomena occurring within the system, researchers

develop and justify new descriptive or explanatory theories, whereas if the goal is to improve the system, they build and evaluate artifacts and corresponding prescriptive design theories ② – ④. The outcomes of

these efforts are both applied to the original system ⑤, and added to the scientific knowledge base for future use ⑥.

Descriptive and explanatory theories provide the understanding needed to design effective artifacts, whereas artifacts embedded in context are the subject of new theories. In the remainder of this article, we will illustrate many of our arguments using design science examples, which, in our opinion, holds the greatest need and the greatest opportunity for advancing the impact of IS on wicked problems. But our arguments hold true for behavioral research as well, and we will highlight several such instances.

<b>Problem Dimension</b>	<b>Methodological Challenge</b>	<b>CB</b> (Power TAC)	<b>TAC</b> (SCM, AA)	<b>RC</b> (Net-flix, Kaggle)	<b>ACE</b>	<b>ABVW</b> (World 3)	<b>Serious Games</b>
<b>Cost of Social Negatives:</b> Failures of real-world interventions, even at small scale, entail prohibitive costs	Reduce negatives through <b>high external validity eval.</b>	+			+	+	+
	Produce rigorous <b>design theories</b>	++	++	++			
<b>Unrealized Challenges:</b> Solutions should preempt anticipated challenges	Produce <b>solutions</b> in addition to <b>insights</b>	++	++	++	+	(+)	+
	<b>Demonstrate viability of candidate interventions</b> for expensive real-world evaluation	++		+	+	(+)	+
<b>Rapid Pace:</b> Real world progresses quickly and unpredictably	Avoid <b>wasteful duplication</b> in developing a joint understanding of the challenge	++	+			++	
	Maintain <b>up-to-date understanding</b> of the challenge	++	+			++	
	Find right <b>abstraction/relevance balance</b>	+			+	+	+
	<b>Benchmark alternative interventions</b> swiftly	++	++	(+)	(+)		(+)
	<b>Disseminate</b> results in a timely manner	++	++	++	+		
<b>Scale and Scope:</b> Wicked problems hve vastly broader scales and scopes that most traditional IS research domains	<b>Interact</b> with all stakeholders	++	(+)			+	
	<b>Understand</b> the problem and <b>find</b> solutions, effectively <b>coordinate</b> many research groups	++	(+)	(+)	(+)	(+)	+
	<b>Evaluate</b> candidate solutions swiftly, rigorously, and with high external validity	++	+		+	(+)	+
<b>Essential Complexity:</b> Increasing use of IS, (smart) markets, and other social forms of organization create essential complexity	<b>Explore broad solution</b> space	++	++	++			+
	<b>Produce comparable artifacts</b> based on shared paradigm	++	++	+		(+)	+
	<b>Comprehensively formalize</b> the problem and solution quality criteria, quickly converge on a research paradigm	++	+				
	<b>Evaluate strategic interactions of evolving candidate solutions</b> , relate locally optimized artifact performance to overall system performance, deal with asymmetrically dominated options	++	++			++	(+)

Table 1: Comparison between Competitive Benchmarking and related methods in terms of ability to resolve key obstacles to IS research on wicked problems. Parentheses indicate that a method, while potentially able to remove a certain obstacle, is usually not used to this end in practice.

This general research framework applies to wicked problems as well as to challenges of smaller scale. However, a number of issues arise in each step of the framework when applying it at the societal level. We discuss these issues below.

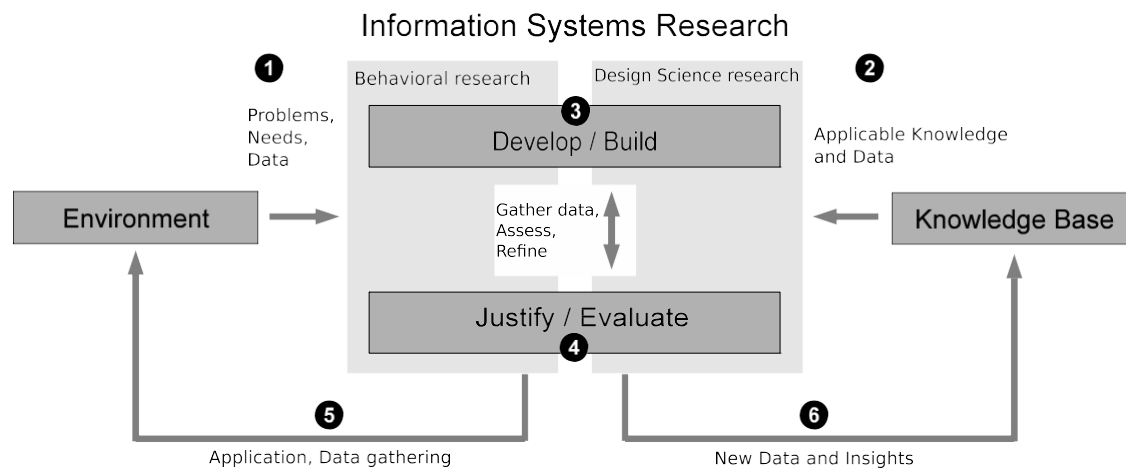


Figure 1: IS research encompasses behavioral and design science research. The depicted framework is adapted from [Hevner et al. \(2004\)](#).

### Defining Problems and Needs ①

Wicked problems exceed the capacity of individual research groups to **interact with all stakeholders** to build and maintain an understanding of an unfolding challenge ([Arias et al., 2000](#)). Bringing different and often controversial points of view together to create a shared understanding among these stakeholders can lead to new insights, new ideas, and new artifacts. For example, a research group attempting to design IT- based interventions to climate change would have to discover, collect, and understand a wide variety of data, interacting with meteorologists, geologists, politicians, chemists, economists, sociologists, industrial and commercial players, and many other stakeholders to develop an understanding of climate change and its expected societal impact. But even if time and resources were unlimited, a wicked problem (such as climate change) **defies comprehensive formalization** of the challenge itself and of the detailed objectives for possible interventions. This is a direct consequence of the essential complexity of the systems such challenges emerge from ([von Hayek, 1989](#)). In the climate change example, an intervention might aim to protect biodiversity, mitigate the short-run impact on the global food supply, or maintain economic growth.

Each of these objectives gives rise to a different set of interventions and to a different delineation of the challenge. In other words, the definition of the challenge, the vocabulary used to describe it, and the questions researchers ask about it all become a crucial part of the challenge itself ([Rittel and Webber, 1973](#)).

Two conventional responses to these issues have been to either work on a small subset of the challenge, or to establish large, centrally composed and hierarchically organized research consortia ([Hey](#)



[et al., 2009](#)). By focusing on small subproblems, researchers ignore essential facets of the challenge, create candidate interventions that cannot easily be compared to interventions for adjacent subproblems, and ignore important system-level consequences. Centrally composed and hierarchically organized research consortia forgo the opportunity of leveraging the diversity of various research groups for understanding the problem from a wide range of angles. Large consortia also tend to move more slowly than the rapidly evolving challenges they aim to address ([Moss et al., 2010](#)). Unsurprisingly therefore, practitioners find “science [to be] lagging behind the commercial world in the ability to infer meaning from data and take action based on that meaning” ([Hey et al., 2009](#)).

We argue that methodological advances are needed to support interdisciplinary communities of stakeholders and researchers in jointly developing (1) problem definitions and models of wicked problems, (2) shared vocabularies, and (3) lists of important research questions. Loosely following [Kuhn \(1996\)](#), we refer to this triplet as a scientific **paradigm**. Any method fit for this purpose must effectively use the limited capacity of individual research groups by facilitating a **separation of concerns** among them. Separation of concerns is a design principle for separating a problem into distinct modules such that each module addresses a separate, clearly defined concern, and that the interfaces or couplings among modules are well-defined and easily understood. In the context of CB, the principle of separation of concerns applies to the division of work among research groups, as well as the design of the platform.

### Using the Knowledge Base and Building Artifacts ② ③

The scale and complexity of societal problems comes paired with a vast number of possible interventions. In our climate change example, these interventions might include organizational redesigns, legislation, economic incentives, deployment of technology, geo-engineering, or a combination thereof. Research on wicked problems must consider a **broad range of diverse candidate interventions** based on experiences of researchers and stakeholders from various disciplines to understand the nature of *good* interventions in the absence of a unique quality criterion ([Pries-Heje and Baskerville, 2008](#); [Collins et al., 2009](#)). In the case of technological interventions, studying a broad range of candidate artifacts is particularly important, because the effects of strategic interactions among artifacts can easily dominate the performance of artifacts studied in isolation ([Hanusch and Pyka, 2007](#)).

Quickly generating and evaluating such diverse candidate interventions presents current scientific methods with difficulties. Lacking a shared paradigm, the current norm tends to produce disparate candidate interventions based on different problem definitions, hampering comparison and improvement.

We argue that methodological advances are needed to foster interdisciplinary communities of researchers working from a shared paradigm. This will require new forms of coordination among many research groups, and a mindset that favors a peer-reviewed and community-owned paradigm over after-the-fact comparisons of results based on disparate problem definitions.

#### **Evaluating Artifacts ④**

Clearly, interventions in complex systems should be evaluated at many levels, including the system level where strategic interaction effects can be observed. This is particularly difficult for societies where the increasing use of markets and other social forms of organization has vastly increased the number and diversity of interactions ([Bichler et al., 2010](#)). Consider the case of the global financial markets with their continuously evolving structures. These markets:

“involve or acquire significant degrees of variability in components and heterogeneity of constituent systems ... For this reason traditional engineering techniques, which are predicated on very different assumptions, cannot necessarily be trusted to deliver acceptable solutions. ... [N]ew approaches are required: new engineering tools and techniques, new management perspectives and practice ([Cliff and Northrop, 2012](#)).”

Formal analysis may provide important insights in stylized settings, but they are necessarily limited when it comes to evaluating complex system interventions ([Tsfatsion, 2006a](#)). Real world evaluations, on the other hand, are problematic because of the prohibitive cost of well-intended interventions gone awry, that is, the **cost of social negatives**. Pilot evaluations could alleviate these risks, but they are expensive and their realism is often bounded by a homogeneous, small-scale setup where one consortium controls the entire pilot. Finally, many important problems like climate change, aging societies, and depletion of carbon-based energy sources have **not fully materialized** yet, rendering real world evaluations simply impossible.

IS researchers have extensive experience with system level evaluations that often include strategizing actors and artifacts, e.g., ([Bapna et al., 2004](#); [Wang and Benbasat, 2005](#)). But because of the **vast number of interactions**, **decentrally evolving artifacts**, and the **evolving web of interactions** among them, interventions in societal challenges are particularly difficult to evaluate. Research must anticipate and preempt societal challenges instead of studying them in retrospect, but it is unclear how researchers can cater to unrealized future needs while meeting standards of academic rigor today.

We argue that methodological advances are needed in system-level evaluations of decentrally evolving artifacts and their strategic interactions for currently unrealized problems of societal scale. Evaluation facilities must provide detailed, comparable data on artifact performance and evolution, and balance swift evaluation against the risk of incurring social negatives. We should emphasize that we do not attempt to prescribe a single best tradeoff between abstraction and relevance. Instead, we see this tradeoff as a conscious choice, jointly made by researchers and stakeholders during the definition of their paradigm.

We have emphasized the need for a broad range of candidate interventions, and we emphatically include casual or ad-hoc designs in this statement. But the end result must be rigorous design theories with high external validity rather than individual, idiosyncratic designs ([Walls et al., 1992](#); [Gregor and Jones, 2007](#)). That is, given the prohibitive costs of social negatives, researchers must strive for prescriptive theories about design rules that work consistently well, under a broad range of conditions, and with high confidence.

### **Communicating with the Environment and the Knowledge Base ⑤ ⑥**

Producing rigorous and impactful results on wicked problems is difficult for at least three reasons. It may be difficult to obtain relevant data about the full scope of the challenge and its environment, and the ways in which various stakeholders are impacted. Stakeholders expect researchers to proactively provide **solutions in addition to insights**. Policy makers, for instance, seek concrete guidance on the technologies, rules, and institutions of future energy infrastructures ([Kassakian and Schmalensee, 2011](#)). Second, due to the scale and complexity of wicked problems it is often difficult to communicate the problem and possible interventions, and to **convince stakeholders of the viability of interventions** for further evaluation in the real world. Finally, the established scientific **publication cycle cannot keep up with the pace of societal challenges**, which reduces the timeliness of research results and their potential impact.

We argue that methodological advances are needed that encourage researchers to produce tangible representations of their results in addition to textual descriptions. These representations must be based on a credible, peer-reviewed paradigm, invite further experimentation by researchers or practitioners, be readily comparable to alternatives, and come with detailed performance records in the form of curated experimental data. By working from a shared paradigm, and by making data and designed artifacts first-class citizens of the scientific process, frictions in building on other researchers' results can be reduced, and the credibility and concreteness of results can be increased.

## Competitive Benchmarking

Competitive Benchmarking (CB) is a novel IS research method that is designed for modeling and evaluating competition-based approaches to wicked problems. At the heart of CB is a separation of concerns around rich representations of scientific paradigms and research results. CB enables scalable interdisciplinary research communities in which coordination and peer review are shifted to the earliest possible time. The return on this up-front investment comes in the form of comparable, actionable research results, and timely dissemination.

The three elements of CB are visualized in Figure 2.

1. CB ALIGNMENT<sup>35</sup> refers to a continuous synchronization process between a scientific paradigm and a wicked problem, and it provides for the timely dissemination of late-breaking results.
2. CB PLATFORM is the medium in which researchers and stakeholders represent an evolving scientific paradigm, and it provides the infrastructure for the PROCESS.
3. CB PROCESS is where independent researchers iteratively build novel theories and design artifacts, while benchmarking and improving their work in direct sight of each other.

In the remainder of this section, we elaborate on these core elements and describe where CB departs from conventional IS research methods.

### Competitive Benchmarking ALIGNMENT

No single research group is likely to understand the full extent of a wicked problem, and we therefore propose a shared scientific paradigm, established through a community-based process. This paradigm must be updated continuously as technologies, regulations, or objectives change (synchronization function). Research results and associated data must be disseminated in a targeted and timely fashion in order to have impact (dissemination function). In CB, these two functions are realized through a continuous ALIGNMENT process. impact (**dissemination** function). In CB, these two functions are realized through a continuous ALIGNMENT process.

---

<sup>35</sup>We use small capitals to distinguish ALIGNMENT, PROCESS, and PLATFORM as defined in Competitive Benchmarking from their usual interpretations.

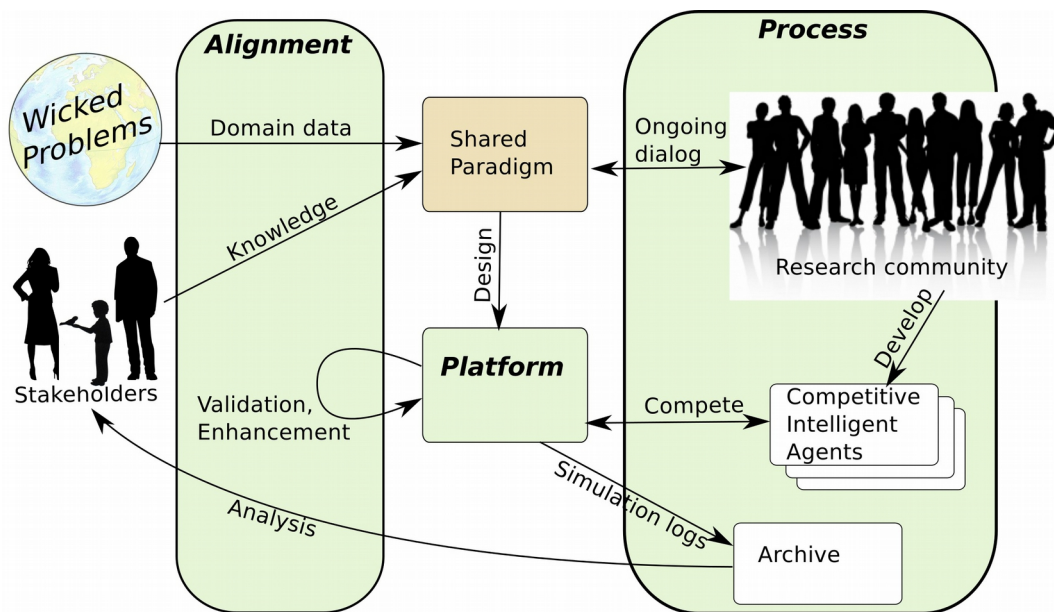


Figure 2: Competitive Benchmarking involves communities of stakeholders and researchers around a shared paradigm and a common platform.

Let us first consider **synchronization**. Establishing and maintaining an accurate model of a wicked problem is an important precondition for research that generates useful theories and artifacts, and that offers reliable policy guidance (Pyka and Fagiolo, 2007). Neither the idea of continuous analysis nor the methods CB researchers use to this end, differ from conventional research and we will therefore not discuss them further (see, e.g., Gray, 2004; Majchrzak and Markus, 2013). ALIGNMENT’s distinguishing feature is that it encourages the establishment of one shared, peer-reviewed paradigm early on, to increase the speed, effectiveness, and credibility of the research efforts that follow.

The basic idea is to replace the single-investigator model and its numerous smaller, incompatible problem definitions with a social learning process that is better suited for gathering and sharing dispersed, often tacit stakeholder knowledge, as well as a body of data that can be used to ground and validate the knowledge base and the resulting models. The resulting paradigm is continuously updated and represented in a software-based CB PLATFORM, a choice of medium that we discuss in detail below. In practice, community-based data gathering and paradigm development requires initial investments from a core community of dedicated researchers. Once a critical mass of groundwork has been laid, its benefits become evident and a virtuous cycle of peer-review, incremental refinement, and increase in paradigm value sets in. As researchers from diverse backgrounds begin adopting and contributing to the paradigm, they increase the community’s capacity for understanding the challenge, improve the coverage and detail of the paradigm, challenge prior assumptions, and provide additional validation.

It is equally important to maintain correspondence between the paradigm and the problem under study. In our own CB efforts, we institutionalize this correspondence through industry and policy **advisory boards**

that meet regularly to provide guidance on important aspects of the problem. The upshot is an intellectual capital base with high managerial and societal relevance, that each researcher is willing to invest in, and that benefits the entire community by providing a high-quality shared research infrastructure.

The goal of ALIGNMENT is not to establish one universally accepted world-view, nor to socialize the scientific process. As we shall see below, CB encourages a type of intense, competitive innovation in which individual achievements are promoted rather than attenuated. But for such competitive innovation to be effective, researchers must start from compatible assumptions and distribute their limited time judiciously. ALIGNMENT provides upfront coordination and open dispute resolution *before* major research efforts are undertaken. It avoids duplicate work during the problem definition phase, it promotes research results that are comparable after the fact, and it leads to a greater confidence that the community's efforts flow into the highest-value research questions.

The results of these efforts must be communicated in a targeted and timely fashion to have impact and to accelerate progress ([Garvin, 1993](#)). CB supports the timely communication of results through the **dissemination** function of ALIGNMENT. Clearly, the community of stakeholders and researchers involved in CB is a natural starting point for dissemination, with a vested interest in results guided by their own ideas. But the dissemination function adds at least two other novel and important benefits.

First, by combining a peer-reviewed paradigm with a swift but rigorous PROCESS, CB offers an alternative to the protracted ex-post review of assumptions and results that is the current scientific norm. A significant share of review is performed up-front at the paradigm level by numerous independent researchers and stakeholders. As pointed out by [Kleindorfer et al. \(1998\)](#), ALIGNMENT is “a way of effecting ... validation. The interaction between the modeler and the client in mutually understanding the model and the process establishes the model's significance; that is, its warranty.” Individual researchers then develop new theories and artifacts based on the validated paradigm, which are ultimately evaluated by an independent party during the public CB PROCESS. There, theories and artifacts have to perform well under demanding conditions that are partly determined by the evaluators, and partly by interaction with other researchers' designs. Fine-grained protocols of these evaluations are made publicly available to support their credibility. Overall, this procedure greatly reduces the need for ex-post scrutiny and time to disseminate.

Second, because ALIGNMENT is problem-centric and continuously seeks to identify the next most important insights and solutions, it reduces the risk of addressing outdated problems. It thereby generalizes the idea of applicability checks ([Rosemann and Vessey, 2008](#)) to a continuous process that guides a research community.

The prohibitive cost of potential social negatives will make decision-makers in industry and policy, understandably, skeptical of trusting just any result. A diligently executed process of ALIGNMENT leads to

an improved rapport with these stakeholders and adds credibility to research results obtained through CB. Combined with timely, tangible results in the form of data and executable artifacts, this creates attractive opportunities for high-impact dissemination.

### **Competitive Benchmarking PLATFORM**

The PLATFORM is the central point of coordination for CB participants. It is the malleable, executable **representation of the shared paradigm** created and updated during ALIGNMENT . It provides a **toolset and access to data for empirical science** to the PROCESS .

Given the central role of the paradigm within CB, the medium used to **represent** it is important. The most common medium, natural language, has three significant shortcomings: it has no safeguards against imprecisions and inconsistencies, it is difficult to update as the problem evolves, and it must often be translated into other media to become actionable. Formal representations address the first concern, but they are limited in terms of problem sizes they can address.

CB instead promotes the use of software-based PLATFORMS and accompanying data that leverage the great strides that software engineering has made in understanding and representing complexity. These started with the realization that modeling complex socio-technical systems should be an iterative, social learning process. Related progress in computer language theory has bred a generation of highly expressive, problem-centric languages that put stakeholder needs before machine considerations ([Meyer, 1998](#)). Advances in program design and architecture have made software extensible and adaptive to changing environments. The upshot is a proven, scalable, and social approach to capturing complexity ([Baetjer, 1997](#)), typically in the form of a simulation model in which one or more competitive entity types are identified and externalized as competitive intelligent agents ([Ketter et al., 2015](#)) that support the competition element of the CB PROCESS . The advantages of software-based paradigm representations come at a greater cost of initially describing the problem at the necessary level of detail, which may need to be spread over several research groups. We also note that technical qualities of software-based representations may require advanced software engineering skills, a point we revisit in the discussion. Among these qualities are a clear design that makes it easy for other researchers to understand, use, and extend the paradigm, good readability and thoroughly documented assumptions, a modular architecture that enables specialist contributions in clearly delineated areas, and a licensing model that encourages free redistribution and extension.

The second PLATFORM function is that of a toolset for empirical science. Because the PLATFORM encodes a shared understanding of a wicked problem, research results and tools derived from it will be comparable and technically compatible. For the purpose of theory validation, PLATFORM data can be

compared to data obtained from studies under different environmental conditions, or reproduced under identical environmental circumstances ([Teshfatsion, 2006b](#); [Pyka and Fagiolo, 2007](#)). Designed artifacts can readily be benchmarked against artifacts from other research groups. Ecosystems of scientific tools can be built around the PLATFORM to aid researchers in routine tasks such as data screening, reporting, and distributed experiment management.

We should emphasize that the presence of an executable representation of the paradigm also means that fully executable interventions like dynamic decision rules, economic mechanisms, or IS artifacts can be built against the PLATFORM. These interventions are tangible and interesting to study for practitioners and researchers alike.

### **Competitive Benchmarking PROCESS**

Any effective research method is a structured approach to exploring and learning about phenomena (descriptive and explanatory research) and solution spaces (design science research). Researchers create new theories and designs, evaluate their realism and usefulness, learn from experience, and iterate to improve their work (see [Figure 1](#)). This structured form of learning and improvement is related to benchmarking in that it requires skills in “systematic problem solving, experimentation with new approaches, learning from ... own experience and past history, learning from the experiences and best practices of others, and transferring knowledge quickly and efficiently.” Its best practitioners “[rely] on the scientific method, rather than guess-work, for diagnosing problems” and “[insist] on data, rather than assumptions, as background for decision making” ([Garvin, 1993](#)).

Suppose a community of researchers and stakeholders is interested in understanding the effects that different transaction tax regimes have on the trading behavior of commercial banks and in stability implications for global financial markets. Starting from these goals, they engage in ALIGNMENT and model the behaviors of private and institutional investors, a market infrastructure, central banks, etc. until they agree on having captured the most salient features of the challenge. The result of this work is an aligned PLATFORM on which the PROCESS proceeds iteratively, each cycle consisting of four phases. [Figure 3](#) visualizes the process, showing the major activities of the community of stakeholders, and of the competition participants, during each cycle.

**Design:** Several research groups design artifacts, typically in the form of autonomous software agents that implement identified competitive entities in the PLATFORM definition. Agent behaviors are typically conditioned offline by a variety of data sources, and online by large amounts of data generated by the PLATFORM. The strategies of these agents can be based on *ad-hoc* designs or on sound kernel theories, as long as they remain within the agreed-upon paradigm.<sup>4</sup> Strategies can even involve human participants,

---

<sup>4</sup>This does not preclude artifacts from exploiting loopholes within the PLATFORM; one of the benefits of CB is the discovery of unintended loopholes through a wide array of creative artifacts.



which opens interesting avenues for work on behavioral theories ([Babb et al., 1966](#); [Collins et al., 2009, 2010](#)). Researchers repeatedly evaluate their strategies against each other and the PLATFORM to detect and remove weaknesses.

**Compete:** Participants then pit their artifacts against each other in a formal tournament where strategic interactions and system-level properties can be observed. An independent party determines the tournament schedule, including the groupings of artifacts and environmental conditions (e.g., physical environment characteristics, tax levels, trading intensities). Environmental conditions can also include “shocks” such as storms and major outages. For participants, good performance in a strong field of competitors is reward and incentive for further improvement.

**Analyze:** The tournament outcome is a ranking of strategies, together with fine-grained data on artifact and system-level behavior. This data is publicly available, its content and format is documented, and tools are provided for extracting interesting subsets. The dataset from a simulation is a complete record, including all inputs, outputs and state changes. It includes seeds for all random sequences to support full reproduction of a simulation scenario. The PLATFORM and its accompanying scientific tools promote credible analyses that can be produced quickly and distributed along with the underlying data.

**Disseminate and Realign:** The insights gleaned from these analyses are disseminated to researchers and stakeholders through formal publications as well as direct interaction with stakeholders. Analyses can, for example, pinpoint drivers of artifact performance that research groups can use to direct their future efforts, e.g., ([Jordan et al., 2007](#); [Ketter et al., 2013c](#)). Researchers also make executable versions of their tournament artifacts available for study, to support empirical research outside the tournament environment. Ongoing discussions with stakeholders and researchers identify issues and priorities to update CB ALIGNMENT for the next cycle in the PROCESS.

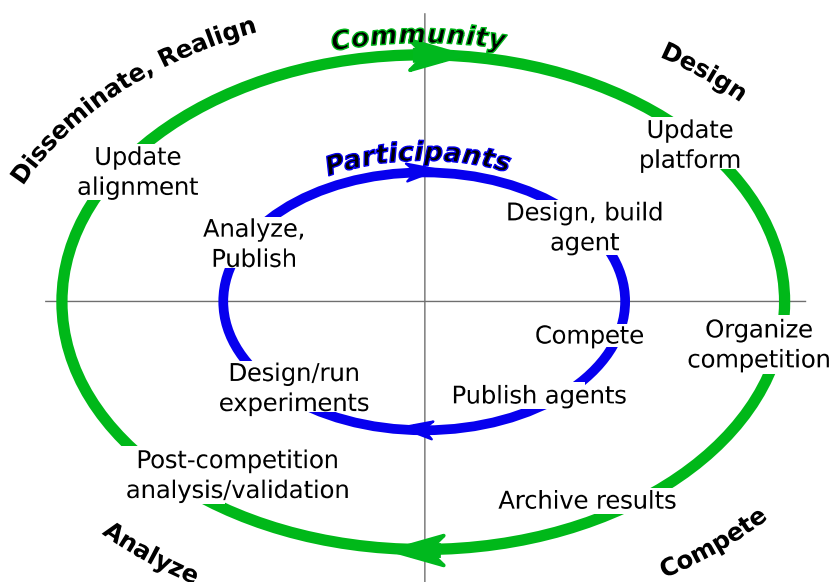


Figure 3: Competitive Benchmarking process cycle.

The CB PROCESS equally supports several types of scientific inquiry that close the IS research cycle described by [Hevner et al. \(2004\)](#). Most importantly, a PLATFORM together with a fixed set of high-performing artifacts can be used as a conventional Agent-based Virtual World (ABVW) to perform controlled experiments in pursuit of descriptive or explanatory theories ([Ketter et al., 2010](#); [Chaturvedi et al., 2011](#)). These theories can then be used by artifact designers to improve their designs. The continuous evolution of artifacts in the PROCESS yields diverse, high-performing artifacts that can be studied towards descriptive or explanatory theories. Examples of supported research types are shown in Table 2.

Research Type	Research Setup	Examples
<b>Artifact Design</b>	<ol style="list-style-type: none"> <li>1. Use PLATFORM for distributed artifact design</li> <li>2. Benchmark and improve artifacts iteratively</li> </ol>	<ul style="list-style-type: none"> <li>- Trading strategies</li> <li>- Dynamic pricing</li> <li>- Brokers</li> </ul>
<b>Controlled Experiments</b>	<ol style="list-style-type: none"> <li>1. Hold set of high-performing artifacts constant</li> <li>2. Execute artifacts against PLATFORM while varying environmental parameters</li> <li>3. Measure resulting system-level properties</li> </ol>	<ul style="list-style-type: none"> <li>- Social welfare studies</li> <li>- Distribution studies                             <ul style="list-style-type: none"> <li>- Concentration and competitiveness measures</li> </ul> </li> </ul>
<b>Falsification Studies</b>	<ol style="list-style-type: none"> <li>1. Vary set of high-performing artifacts</li> <li>2. Execute artifacts against the PLATFORM</li> <li>3. Assess stability of mechanism or theory</li> </ol>	<ul style="list-style-type: none"> <li>- Market mechanisms</li> <li>- Circuit breakers</li> </ul>
<b>Mixed Initiative Studies</b>	<ol style="list-style-type: none"> <li>1. Vary set of high-performing artifacts</li> <li>2. Human participants</li> <li>3. Execute artifacts against PLATFORM</li> <li>4. Assess human or artifact performance</li> </ol>	<ul style="list-style-type: none"> <li>- Decision support systems</li> <li>- User interfaces</li> </ul>

Table 2: CB supports descriptive and explanatory research (insights) as well as design science research (solutions).

CB's PROCESS contains four novelties that aim to improve the capacity of IS research for tackling wicked problems. Most importantly, it adds **naturalistic dynamics** to artifact validations. In our example, researchers cannot hope to experiment with real tax regimes and must therefore resort to working against a model of the challenge ([Smith, 1982](#)). However, one particularly important facet of real-world evaluation *can* be brought into the laboratory: the competitive co-evolution of artifacts. Like firms and individuals in the real world, CB participants constantly seek to improve their designs by adapting to the behavior of the environment and of others in a type of Emergent Knowledge Process (EKP, [Markus et al. 2002](#)). The

ensuing dynamics provide a unique tradeoff between artificial and naturalistic elements for high-risk evaluations in complex economic environments.<sup>5</sup>

Second, the aligned PLATFORM is **validated by other researchers and stakeholders**, and evaluation conditions are determined by an independent party. That artifacts and theories must perform well under many different circumstances in a realistic environment increases **external validity** and researchers' confidence in the absence of unanticipated social negatives.

Third, community-based ALIGNMENT and PLATFORM development spreads the effort of understanding and modeling a challenge across many researchers to **increase scientific cycle speed**. The initial investment amortizes as researchers gain the ability to rapidly test artifacts and theories without the frictions of first finding compatible benchmarks. Publicly evaluated artifacts and theories can then be swiftly disseminated. Evaluation data also can be used to derive **rigorous design theories**, which is an important step in reconciling the need for scientific rigor with leveraging the creativity of pragmatic designs. It may not even be known why a particular artifact works at the time of evaluation, but the availability of evaluation data allows the community to discover theoretical principles behind its working later on.<sup>6</sup>

And finally, the comprehensive data generated in the PROCESS provides **clear visibility of the progress** that designers make in improving their artifacts, which also gives a measure of the benefits of CB as a research method ([Venable and Baskerville, 2012](#)). When progress tapers off, the community may also decide to call its advisory board for new challenges.

### **Interaction effects of ALIGNMENT , PLATFORM , and PROCESS**

We should emphasize that CB does not attempt to replace the existing process of scientific knowledge discovery. It rather aims to remove several common obstacles, and it adds a structured approach to benchmarking which, in our opinion, is insufficiently represented in current IS research practices. One of the resulting benefits for IS research on wicked problems is a clear separation of concerns between various stakeholder and researcher groups around the PLATFORM , which ultimately leads to better scalability, and which we summarize in Table 3.

---

<sup>5</sup>An alternative view on this is based on the “increasing recognition of the mutable nature of these artifacts. That is, they are artifacts that are in an almost constant state of change” ([Gregor and Jones, 2007](#)). Designs, in the context of CB, are by definition “evolutionary trajectories,” not static blueprints, and an important benefit of CB is the ability to generate such trajectories realistically, and to study their development over time.

<sup>6</sup>A similar separation of concerns led Johannes Kepler to discover the laws of planetary motion from recordings in the notebooks of Tycho Brahe ([Hey et al., 2009](#)). We speculate that the lack of comparability between artifacts causes this separation of concerns to be virtually absent from design research today.

<b>Separation of concerns between</b>	<b>Enables ...</b>
<b>Stakeholders and Researchers</b>	<ul style="list-style-type: none"> <li>- Researchers to effectively learn about the challenge</li> <li>- Stakeholders to learn about new research insights and solutions in a timely fashion</li> </ul>
<b>Researchers</b> from different disciplines or with different expertise	<ul style="list-style-type: none"> <li>- Scalable, expert model-building and concurrent work on one joint problem definition. For example, a battery expert might build realistic models of e-vehicle charging behavior to be used by an economist in the design of market mechanisms.</li> <li>- Competitive design. For example, a machine learning (ML) expert and an operations research (OR) expert might design alternative solutions to a given problem. The shared use of a PLATFORM ensures that their artifacts remain technically compatible and comparable.</li> </ul>
<b>Theory/Artifact Designers and Data Scientists</b>	<ul style="list-style-type: none"> <li>- Independent data analysis and validation. PROCESSES generate publicly available data for analysis. An economist could, e.g., analyze the welfare effects of deploying the ML- and OR-based artifacts described above.</li> </ul>
<b>Academic Researchers and Pragmatic Designers</b>	<ul style="list-style-type: none"> <li>- Leveraging the creativity of pragmatic designers (<a href="#">Hevner and Chatterjee, 2010</a>). CB imposes very few constraints on the theoretic underpinnings of designed artifacts. Practitioners can contribute high performing ad-hoc artifacts that are then further analyzed by academic researchers.</li> <li>- Effective industry cooperations. Industrial designers can contribute artifacts that are rigorously evaluated according to the standards of design theories.</li> </ul>

Table 3: CB’s three core elements facilitate an effective collaboration between various groups of contributors. This separation of concerns leads to better scalability in the challenge size, and in the number of independent contributors.

The improvement in scalability stems partly from reducing the waste and redundancy inherent in incomparable research results, and partly from redistributing efforts between individuals and the community. In particular, the early coordination during ALIGNMENT enables the reuse of domain knowledge obtained from stakeholders, and of the scientific toolset provided by the PLATFORM. In other words, individual effort is supplemented by community effort in defining the problem and in evaluating and communicating results. The upshot is more time spent on the value-generating core activities of theory development and artifact building for each individual researcher.

The next section shows the principles of CB at work within a concrete research effort on sustainable energy systems that we have been conducting together with a global community of researchers over the past five years.

## **Power TAC: Data-driven Competitive Benchmarking for Sustainable Energy Systems**

From relatively modest beginnings 130 years ago, electricity has revolutionized the way we live our lives and organize our societies. Unfortunately, the economic benefits are increasingly offset by environmental and sustainability concerns. The drivers behind these negatives are numerous and complex, but one important underlying theme is the mismatch between increasing demands for volume, sustainability, and affordability on one hand, and hierarchical control structures that are largely unchanged from electricity's early days on the other.

Modernizing these control structures is an extremely challenging proposition. The **Smart Grid** ([Amin, 2002](#)) of the future will have to (a) efficiently allocate electricity among hundreds of millions of users with unique preferences, (b) integrate production from renewable and decentralized power sources like rooftop solar panels, (c) respect complicated constraints imposed by grid topology, power flow physics, privacy concerns, and several layers of regulation, and (d) uphold real-time control under uncertainty, all the while ensuring a smooth transition from the operational grid of today. IS scholars can make substantial contributions to this grand wicked problems by “integrating new information and communications technologies, combining them with active support from electricity consumers, and leveraging the optimizing power of markets” ([Coll-Mayor et al., 2007](#)).

The scale and complexity of the problem, and the interrelated advances required in theory and artifact design prompted a global community of researchers to address it with CB through the **Power Trading Agent Competition (Power TAC)**, ([Ketter et al. 2013b](#); [2013c](#), see also [www.powertac.org](http://www.powertac.org)). Power TAC fills several recently proposed IS research agendas on energy and sustainability ([Bichler et al., 2010](#); [Melville, 2010](#); [Watson et al., 2010](#)).

### **Power TAC ALIGNMENT**

The idea for Power TAC originated in 2009 during a workshop with stakeholders from German government, science, and industry. The Power TAC project began later in 2009 with a core group of researchers, who surveyed the literature on power systems, smart grid concepts, and sustainability issues, gathered data from a variety of sources, and interviewed stakeholders to develop an initial ALIGNMENT . Key stakeholders were identified in utility companies, network infrastructure providers, communication electronics manufacturers, electricity cooperatives, public policy, and electricity customer lobby groups. Key data included records of wholesale market activity in several jurisdictions across Europe and North

America, weather and weather forecasts from areas covered by the wholesale markets, detailed records of household energy consumption from multiple pilot studies, terms and conditions of published tariffs from areas with retail competition, data on driving patterns and charging behavior for electric vehicles in Europe, and a variety of other sources. Stakeholders were interviewed repeatedly, and many joined an advisory board which now institutionalizes Power TAC’s ongoing ALIGNMENT. The board meets periodically to provide researchers with industry insights, to ensure that important problems are being tackled, and to disseminate the latest research results.

Table 4 shows a sampling of data resources that have been used for Power TAC ALIGNMENT. This data is used for various purposes, such as modeling different user and power types, and simulating their behavior afterwards under different scenarios. For example, the car2go data has been used to model an electric vehicle fleet (Kahlen and Ketter, 2015) that provides grid-stabilization services in addition to mobility services.

Data Type	Institution	Source
Wholesale Market	Australian Energy Market Operator	<a href="http://www.aemo.com.au">www.aemo.com.au</a>
Wholesale market	European Energy Exchange	<a href="http://eex.com">eex.com</a>
Wholesale market	Midwest ISO	<a href="http://www.misoenergy.org">www.misoenergy.org</a>
Wholesale market	Ontario’s IESO	<a href="http://www.ieso.ca">www.ieso.ca</a>
Demographics & Mobility	Dutch Statistics Office	<a href="http://statline.cbs.nl">statline.cbs.nl</a>
Demographics & Mobility	German Statistics Office	<a href="http://www.destatis.de">www.destatis.de</a>
Demographics & Mobility	car2go	<a href="http://code.google.com/p/car2go">code.google.com/p/car2go</a>
Smart Grid pilot project	Pecan Street Project	<a href="http://www.pecanstreet.org">www.pecanstreet.org</a>
General	US Energy Information Agency	<a href="http://www.eia.gov">www.eia.gov</a>

Table 4: Samples of publicly available data used for Power TAC ALIGNMENT.

After several ALIGNMENT iterations, Power TAC began to attract outside researchers interested in leveraging the publicly available PLATFORM for their own work. Several groups contributed specialized knowledge that improved its realism in areas where no other community member possessed the requisite expertise or resources, e.g., customer modeling (Reddy and Veloso, 2012) and balancing (de Weerd et al., 2011). In addition, work continued to generate and gather relevant data; for example, Koroleva et al. (2014) constructed and ran a social-media experiment to gather data about electric-vehicle charging preferences. In exchange, the contributors could study their models in a rich, realistic environment that they could not have created otherwise, including a dedicated community that validated and critiqued their models. Other groups created experimental tools and third-party analyses of Power TAC (Babic and Podobnik, 2013; Kahlen et al., 2012), compared the PLATFORM against real-world behaviors (Nanoha, 2013), and designed and evaluated artifacts, e.g., (Peters et al., 2013; Kuate et al., 2013; Urieli and Stone, 2014b). Importantly, many of these new participants had technical expertise but no prior domain knowledge or interest in contributing to the sustainable energy problem. It was the availability of a community-supported, executable model of a real-world problem and a list of important research questions that triggered them to apply their

diverse technical skills to sustainable energy. Conversely, researchers and external stakeholders with energy domain knowledge benefited from the innovative contributions of these technical experts.

Our example illustrates how ALIGNMENT provides scalability to communities of researchers coordinating through a shared paradigm. Establishing and maintaining this paradigm regularly requires incisive modeling decisions from the community. But through ongoing ALIGNMENT, these decisions can be made early, thereby keeping subsequent research results technically and conceptually comparable. For example, Power TAC currently:

- models the electric **distribution system** but not the transmission system, because while controlling the latter is well understood, much scientific guidance is needed on making the former “smarter” ([EPRI - Electric Power Research Institute, 2011](#)).<sup>7</sup>
- models the **economic aspects** of the smart grid but not the physical power flows, because of an urgent need for insights on how a combination of IT and economic forces can incentivize sustainable electricity consumption patterns ([Watson et al., 2010](#)).
- models **retail electricity tariffs**, but (so far) not bilateral price negotiations with commercial customers, because end users “can provide remarkable local intelligence ... [but] any technology is doomed to fail if the involved users do not like or understand it” ([Palensky and Dietrich, 2011](#)).

These ALIGNMENT results are continuously translated into the executable and peer-reviewed Power TAC PLATFORM.

### **The Power TAC PLATFORM**

The PLATFORM models a competitive retail power market in a medium-sized city, in which consumers and small-scale producers may choose from among a set of alternative electricity providers, represented by competing **Brokers**. Brokers are autonomous software agents, built by individual research groups. The remainder of the paradigm is modeled by the PLATFORM visualized in Figure 4. The individual models within the PLATFORM are either derived from or driven by data. For example, customer model behaviors are derived from statistical analysis of a large smart-grid pilot project in E.U., and weather data is actual historical observations and forecasts from multiple locations in North America and Europe. The E.U. pilot data is proprietary, so we are limited to using a statistical approach for the public platform. Brokers offer electricity

---

<sup>7</sup>The distribution system is responsible for providing regional electricity to commercial and residential end-customers. The transmission system is where large-scale generators like wind farms and coal power plants feed in high-voltage electricity for long range transmission.

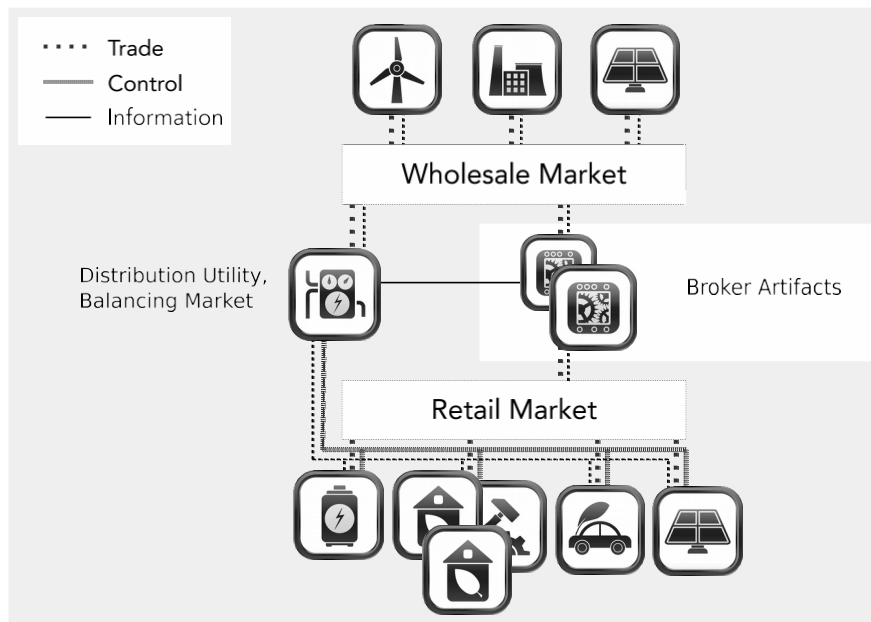


Figure 4: Main elements of the Power TAC paradigm. Brokers are autonomous software agents built by individual research groups. The remainder of the scenario is modeled by the PLATFORM

tariffs (also known as plans or rates) to household and business customers through a retail market. Some customers are equipped with solar panels and wind turbines, which produce and consume power, and many own demand-side management capabilities such as remotely controllable heat pumps or water heaters. All customers are equipped with smart meters from which consumption and production is reported every hour. Customers are sensitive to price changes, weather conditions, and calendar factors such as day of week and hour of day, and they have a range of preferences over tariff terms. For example, some are willing to subscribe to variable-rate tariffs if they have the opportunity to save by adjusting their power usage, while others are willing to pay higher prices for the simplicity of fixed-rate or time-of-use tariffs. Many of these models are contributions from the user community, e.g., ([Gottwalt et al., 2011](#); [Reddy and Veloso, 2012](#)). Brokers buy and sell energy from retail customers and a day-ahead wholesale market, where utility-scale power suppliers sell their output. These suppliers represent different price points and lead-time requirements, e.g., fossil and nuclear power plants, gas turbines, and wind parks.

The **Distribution Utility** (DU) models a regulated monopoly that owns and operates the physical facilities (feeder lines, transformers, etc.) and is responsible for real-time balancing of supply and demand within the distribution network.<sup>810</sup> It does this primarily by operating a balancing market, the real-time facet of the wholesale market, and by exercising demand and supply controls provided by Brokers. The associated costs are allocated to imbalanced Brokers. Given a portfolio of customers, Brokers compete in the wholesale market to minimize the cost of power they deliver to their consuming customers, and to maximize the value of power delivered to them by their producing customers.

<sup>810</sup>In the real world, balancing responsibility is typically handled at the transmission level; the simulation implements a generalization of proposals to move some balancing responsibility to the distribution level ([Strbac, 2008](#)).



The Power TAC PLATFORM as described here has quickly evolved into a very comprehensive economic simulation for smart distribution networks worldwide. Its source code is licensed under a research and business-friendly Apache license and can be freely downloaded from <https://github.com/powertac>. Use and modification of the PLATFORM are not restricted to those participating in the PROCESS, but several important CB benefits can only be reaped by actively engaging with the Power TAC community in this way.

## **The Power TAC PROCESS**

Power TAC uses the CB PROCESS to set a pace and keep a number of research groups around the world engaged, coordinated, and productive. The process began with a three-year focus on ALIGNMENT and PLATFORM-building, starting with a workshop in 2009 sponsored by the German government. Much of the ALIGNMENT effort was in finding, analyzing, and understanding data from a variety of sources, including:

- Energy production, and the costs and relevant physical attributes of various production resource types;
- Characteristics and costs of resources for maintaining grid stability.
  - Wholesale and retail markets for energy, production and regulating capacity, and transmission/distribution capacity;
- Electricity consumption patterns of household, business, and institutional customers;
  - Weather observations and forecasts, and interactions between weather and energy production and consumption.

We used the resulting data and analyses as input for constructing various elements of the Power TAC PLATFORM. For example, analysis of household energy use by [Gottwalt et al. \(2011\)](#) was used to construct customer models, including the “factored customer” model by [Reddy and Veloso \(2012\)](#). Weather data, in the form of actual weather reports and forecasts, is used directly in the PLATFORM to drive various consumption and production behaviors, which in turn affects demand and price in the markets.

By the middle of 2011, we had an initial specification ([Ketter et al., 2011](#)) and a working simulation. This enabled us to initiate the next phase of the PROCESS: recruiting other research groups to participate by building their own retail broker agents to compete in the Power TAC markets, and by testing and critiquing the Power TAC PLATFORM. To be competitive, broker agents must analyze and respond to a continuous stream of data from the PLATFORM. Their behaviors are commonly based on machine learning techniques, trained on analyses of past tournaments as well as real-world market data ([Peters et al., 2013](#); [Urieli and Stone, 2014a](#)). We ran several trial competitions in conjunction with international conferences, including IJCAI 2011 in Barcelona, at AAMAS 2012 in Valencia, and at IEEE SG-TEP 2012 in Nuremberg. Starting

in 2013, the Power TAC community has held annual championships at AAAI 2013 in Bellevue, WA, at AAMAS 2014 in Paris, and at AAMAS 2015 in Istanbul.<sup>9</sup>

Tournament scenarios typically model about two months in the simulation environment; each takes about two hours to run. Each tournament begins with a qualifying round of 8-12 days during which Brokers are screened for technical flaws and communication failures, followed by a final round that typically runs 4-6 days. Each round consists of some number of “sets” of simulation runs, each of which includes all combinations of  $n$  brokers taken  $m$  at a time, where  $n$  is the number of brokers competing in the tournament, and values for  $m$  are chosen to provide a range of competitive environments for the brokers. For example, in the 2015 tournament used combinations of 3, 9, and 11 brokers. This design allows us to study the effect of the competitive environment on brokers, markets, and customers, and places a premium on the ability of broker agents to adapt their behaviors to the level of competition. Table 5 lists all finalists in the 2012 Nuremberg pilot and the 2013, 2014, and 2015 Power TAC championships. These Brokers were designed by researchers with expertise in Artificial Intelligence, Electrical Engineering, Information Systems, Machine Learning, and other areas, and their heterogeneous design approaches have contributed to a rich repository of design ideas, executable artifacts, and artifact performance data. All the data from the championship tournaments are publicly available for analysis, as are the specific versions of the Power TAC PLATFORM and executable copies of most of the agents that ran in these tournaments.

<b>Broker</b>	<b>Institute</b>	<b>Country</b>
AgentUDE	University Duisburg-Essen	Germany
AstonTAC	Aston University Birmingham	UK
COLDPower	INAOE, Natl. Institute for Astrophysics, Optics, and Electronics	Mexico
CUHKTac	The Chinese University of Hong Kong	China
CrocodileAgent	University of Zagreb	Croatia
cwiBroker	CWI, Natl. Research Institute for Mathematics and Computer Science	Netherlands
LARGE	Erasmus University Rotterdam	Netherlands
Maxon	Westfälische Hochschule	Germany
Mertacor	Aristotle University Thessaloniki	Greece
MinerTA	University of Texas at El Paso	USA
MLLBroker	University of Freiburg	Germany
NTUTacAgent	Nanyang Technological University	China
Sharpy	Hebrew University of Jerusalem	Israel
SotonPower	University of Southampton	UK
SPOT	University of Texas at El Paso/NMSU	USA
TacTex	University of Texas at Austin	USA

Table 5: Participants in the 2012-2015 Power TAC finals. The list excludes several other participating groups who did not qualify for the final rounds.

<sup>9</sup>AAAI = Association for the Advancement of Artificial Intelligence; AAMAS = Autonomous Agents and Multiagent Systems; SG-TEP = IEEE Conference on Smart Grid Technology, Economics, and Policies.

It is tempting to attempt to compare performance across tournaments, but this ignores the fact that the Power TAC PROCESS includes an additional ALIGNMENT cycle every year, which generally results in updates and additions to the PLATFORM. As a consequence, the results from year to year are not strictly comparable. However, this does not preclude researchers from doing further empirical studies beyond what is supported by data from the annual tournaments. The Power TAC PLATFORM is open source, and the versions used in each year's tournament are documented and archived. Logs from the tournament games contain the full configuration information, weather data, prices, and seeds for all random-number generators used by the various models. In addition, tournament participants are asked to make executable versions of their agents available at the conclusion of each tournament. Given a few machines with network connections, the PLATFORM and agents can be used by anyone for their own purposes, unconstrained by the practical requirements of running an international tournament. Simulations can run for days or weeks if desired, customer models can be reconfigured, new models can be introduced, and the agents can be reconfigured or modified.

We summarize in the following sections a few of the most important changes to the Power TAC PLATFORM between the 2012 pilot competition and the three championship tournaments in 2013-2015. Many of these changes are driven by stakeholder interaction focused on modeling and evaluating policy options and implications.

**2012-2013** brought changes in customer tariff evaluation, a change in balancing market pricing to improve incentives, the ability for brokers to offer variable-rate tariffs, and a change in wholesale market settlement from time-of-trade to time-of-delivery.

**2013-2014** saw introduction of thermal and battery storage customer types, along with ability to pay customers for exercise of regulation capacity. Prior to the 2013 version, brokers could offer the ability to curtail consumption of certain customers to the balancing market as up-regulation capacity, but the customer could be compensated only by an overall discount on energy prices. The resulting complexity and uncertainty discouraged broker developers from using this feature. We also added a cold-storage warehouse model that includes substantial thermal storage capacity and a control mechanism that supports both up-regulation and down-regulation. This allowed brokers for the first time to offer the full range of regulation capacity to the balancing market. Discussions among stakeholders, along with exploration of price records from multiple wholesale markets in North America and Europe showed that the supply curve in the simulator prior to 2013 was a poor approximation to real-world pricing; as a result, we re-designed the

wholesale supplier model to more accurately reflect real-world pricing, and we added a minimum order quantity requirement for orders in the wholesale market. Finally, we corrected an error in the handling of revoked tariffs, in which a specially-crafted tariff could get some customers to pay a withdrawal fee when the broker revoked the tariff.

**2014-2015** introduced three new customer models:

1. An electric vehicle (EV) model is based on statistics of Dutch driving behavior ([Valogianni et al., 2013](#)), segmented by demographic and socio-economic categories. The EV model is the first customer type that is not always connected to the grid. When it is connected, it may offer both up-regulation and down-regulation, including vehicle-to-grid capacity, within the constraints of the customer's need for battery capacity to support driving needs.
2. Electric forklift trucks are managed as fleets within warehouse environments. Each fleet operates according to a weekly shift schedule, and any surplus charging capacity can be used by brokers as regulating capacity. Because its schedules are not affected by weather and are known in advance, the forklift-truck model is able to optimally exploit time-of-use and variable-rate tariffs to minimize its costs.
3. A "solar leasing" producer models a population of rooftop-solar installations with sufficient capacity to strongly affect wholesale prices when the sun is shining.

**2015-2016** will introduce peak-demand pricing for brokers to create incentives to manage demand peaks. This is an important feature of real-world electricity markets that is so far not modeled in Power TAC, partly because the time periods modeled by the tournament scenario are much shorter than the periods over which peak-demand charges are assessed in most real-world jurisdictions. At the time of this writing, we are conducting a discussion among stakeholders on alternative models that will create the needed incentives without penalizing brokers for increasing their market shares. We also intend to introduce EV fleets that are leased to drivers on hourly and daily terms, based on a large body of data from fleets in three German cities and San Diego, California ([Kahlen and Ketter, 2015](#)), as well as a model of parking structures that offer EV charging ([Babic et al., 2015](#)). The 2016 annual competition will be held in conjunction with the International Joint Conference on Artificial Intelligence (IJCAI-2016) in New York.

## Discussion

Throughout this article we have portrayed Competitive Benchmarking as an effective combination of existing tools and techniques, integrated into a coherent method for IS research on wicked problems. In this section, we discuss connections with several influential streams of work, and we describe a set of best practices for implementing CB based on our own experiences with Power TAC.

## Impact

Since 2011, papers published as part of the Power TAC project have generated over 150 **citations** by authors who are not part of the core project group. For example, [Hernandez-Leal et al. \(2015\)](#) describe a novel bidding strategy for a wholesale energy market that constantly adapts to a mix of non-stationary opponent behaviors. [Bae et al. \(2014\)](#) discusses the promise of retail competition for electric service, and the need to model and explore a variety of business and pricing models to achieve individual and societal benefits.

Heightened awareness of sustainability challenges and opportunities in tackling them using large-scale data analytics and decision-making have led to establishment of **new courses** (“Energy Analytics”, “Energy Information Systems”, and “Analytics for Sustainability”) at the graduate level at leading universities around the world<sup>10</sup>. In addition, dozens of master and PhD theses on sustainable business models using data analytics based on the Power TAC project and real-world energy data.

A unique combination of stakeholders including energy practitioners, policy makers, and researchers from economics, computer science, and behavioral science are brought together by the Erasmus Centre for Future Energy Business ([www.erim.eur.nl/centres/future-energy-business](http://www.erim.eur.nl/centres/future-energy-business)), an interdisciplinary energy analytics and market research center at Rotterdam School of Management, Erasmus University (RSM). The center's annual Erasmus Energy Forum ([www.rsm.nl/ef](http://www.rsm.nl/ef)), focuses on energy analytics, new business models, policies, and progress on Power TAC and related projects.

The Power TAC experience has led to additional projects including the EU project Cassandra-Energy ([www.cassandra-fp7.eu](http://www.cassandra-fp7.eu)), which used Power TAC platform in real pilot experiments for strategic decision-making. In a different project, we are using Power TAC as a testbed for large-scale smart sustainable energy cooperatives linking ports and cities. Grants totaling around 4.5 million Euro have been awarded from EU, the Siebel Energy Institute, and several companies.

---

<sup>10</sup>such as, Rotterdam School of Management at Erasmus University, Stern School of Business at New York University, and Haas School of Business at UC Berkeley.

We have been approached by **governments** to use Power TAC to study several possible future scenarios. We are currently developing scenarios for the Port of Rotterdam, which creates annually about 19% of the CO<sub>2</sub> emissions of the Netherlands. The goal of the project is to cut them in half by 2025.

## **Best Practices**

Power TAC is the first comprehensive implementation of the ideas presented in this article. In this section, we present a set of best practices that we have collected so far, and that can inform future CB researchers. CB's reliance on software-based representations is perhaps the most visible departure from established IS research methods. These representations have a fruitful tradition in the software engineering field, and agent-based models in particular have long been accepted into the scientific mainstream ([Gilbert and Troitzsch, 2005](#)). But a scientific paradigm must be readily understandable and modifiable before it is accepted by a community. Purely code-based representations become problematic as the diversity of the community grows and we must therefore emphasize the important distinction between **software** (which includes logical representations like decision trees and process charts) on the one hand, and **program code** on the other. For a PLATFORM to successfully represent a scientific paradigm requires high-quality software, not just high-quality program code. We suggest that PLATFORMs that consist only of the latter are likely to fail in addressing interdisciplinary challenges of societal scale, and we recommend a clear distinction between logical models of the paradigm, and their translation into machine-executable statements. While the former are of critical importance to CB, the latter can be created by experienced engineers who do not necessarily have to be involved in the core research effort.

The intention behind this recommendation is not to downplay the difficulty or importance of actually building a PLATFORM. In CB, as in software projects generally, it is easy to underestimate the effort and skill needed for the implementation of an idea. A PLATFORM must provide a sound foundation on which the PROCESS can proceed, which in turn requires technical qualities like easy installation, configurability, and extensibility. Moreover, the distributed nature of ALIGNMENT requires an understanding of concepts like change control and configuration management that are well-understood by professional software engineers but not often by amateur programmers. In Power TAC, many of the initial software engineering tasks were done by experienced programmers that are part of our research community. But since these initial days, we have gradually hired several full-time software engineers who are now responsible for maintaining the technical foundation of Power TAC's PLATFORM. Separating the logical problem definition from its translation into code is another important separation of concerns and we recommend this also because it frees up additional research capacity.

Moving to the conceptual level, one critique we sometimes encounter is that formal representations like those built on by CB require a full understanding of the dynamics of each part of the system, some of which may be unknown for wicked problems that have yet to unfold. For example, it is largely unclear how electricity consumers would react to real-time prices, simply because such pricing schemes are technically unfeasible or forbidden in most states. While we consider the critique legitimate, we feel that forcing modelers to make all assumptions explicit is a benefit of agent-based modeling and, by extension, CB. Perhaps a more fitting critique is that CB requires the right choice of abstraction. We might counter that artificial systems such as markets or other social forms of organization exhibit:

“properties that make them particularly susceptible to simulation via simplified models. ... [T]he possibility of building a mathematical theory of a system or of simulating that system does not depend on having an adequate micro theory of the natural laws that govern the system components. Such a micro theory might indeed be simply irrelevant” (Simon, 1996, p.19).

But the combination of naturalistic and simulation-based elements in CB indisputably inherits strengths and weaknesses from each that researchers must carefully consider before embarking on a CB effort (e.g., [North and Macal 2007](#)). In particular, skillful modeling remains as critical in CB as it is in any other simulation effort. One common modeling pitfall is the temptation to *boil the ocean*, or attempt to capture every possible detail. ALIGNMENT and PROCESS are purposely iterative and allow researchers to start small and gradually increase the level of sophistication.

Our previous recommendation comes with one caveat: PLATFORMS must provide a certain level of realism before they attract a research community. This should not keep CB initiators from iteratively building their understanding of a wicked problem and the corresponding PLATFORM. But they must brace themselves for an initial investment of time and resources for which current academic incentive systems offer little reward. In the case of Power TAC, it took approximately two years for the PLATFORM to become attractive and stable enough for other researchers to build upon it. We advise future CB researchers to carefully plan this period, and to ensure that resources are available for its duration.

Finally, we should remark that some challenges stand to gain more from CB research than others. In our opinion, these are the challenges that are large in scale and scope, characterized by essential complexity and prohibitive costs of potential social negatives, and that require interrelated advances in theory development and design. Many important CB principles certainly carry over to, for example, common design problems like those currently tackled in research competitions. But CB’s characteristic up-front investments in ALIGNMENT and PLATFORM development offer the highest benefits in situations

where the swift, community-based development of a shared paradigm, the strategic co-evolution of artifacts, and system-level evaluations matter.

## Conclusions

Many important challenges of our time are “wicked problems” that transcend individuals, organizations, and markets, which have been the traditional focus areas of IS research. Problems such as sustainable energy, climate change, and financial market stability can only be fully understood through discovery and analysis of a wide variety of data at many levels of detail. Such problems require interrelated advances in data discovery and analysis, theory development, and design that are best provided by interdisciplinary research communities ([European Commission, 2011](#)). IS innovations have fueled these challenges through their enabling role in globalization, and they should play a similarly important part in their resolution.

Any intervention in complex social systems requires careful consideration of system-level consequences including potential social negatives ([Rittel and Webber, 1973](#)). This is a strong argument for simulation-based approaches ([Smith, 1982](#)). We argue that the single-investigator model of IS research is limited in its ability to scale to the societal level, and to deliver proactive solutions in addition to reactive insights. Competitive Benchmarking effectively addresses these limitations through a coherent combination of ideas from Big Data Analytics, Benchmarking, Trading Agents research, Agent-Based Computational Economics, Agent-Based Virtual Worlds, and several other fields. CB scales to large, interdisciplinary communities of researchers, and it encompasses both behavioral research (insights) and design science (solutions).

At the heart of CB is the notion of a community-created problem definition that we call a *scientific paradigm*. This paradigm includes a software-based PLATFORM along with a rich body of data that grounds and validates the representation. The paradigm and PLATFORM serve as the foundation for a competitive research PROCESS in which artifacts and theories are created, benchmarked, disseminated, and iteratively improved. This PROCESS starts from a set of peer-reviewed assumptions and curated data sources that breeds artifacts and theories, and proceeds through a cycle of design, competition, analysis, dissemination, and realignment. Ongoing interactions among researchers and stakeholders are used to identify important issues, set priorities, and drive policy discussions. The shared paradigm reduces the need for protracted ex-post scrutiny and increases scientific cycle speed.

Power TAC is a concrete instance of CB that addresses the sustainable energy challenge. Over the last five years, the community has created a diverse set of candidate designs for a novel class of IS artifacts



that we call Brokers, that can contribute to sustainability objectives like better integration of renewable energy sources through market-based incentives.

CB as a research method is itself a designed artifact, and the Power TAC process is one example of how CB's comprehensive data archives and agent repositories reach beyond analyses of theories and designed artifacts, into the assessment of the method itself. This has important ramifications as this new level of visibility allows the Power TAC community to purposely control the degree of novelty and challenge admitted into the PROCESS, and it provides a sound measurement of the rates of insight and innovation delivered ([Venable and Baskerville, 2012](#)).

## References

- Amin, M. March 2002. "Restructuring the electric enterprise: Simulating the evolution of the electric power industry with intelligent adaptive agents", In Faruqui, A. and Eakin, K., editors, *Market Analysis and Resource Management*, chapter 3. Kluwer Publishers.
- Andrews, J., Benisch, M., Sardinha, A., and Sadeh, N. 2009. "Using information gain to analyze and fine tune the performance of supply chain trading agents", In *Agent-Mediated Electronic Commerce and Trading Agent Design and Analysis*, pp. 182–199. Springer.
- Arias, E., Eden, H., Fischer, G., Gorman, A., and Scharff, E. 2000. "Transcending the individual human mind – creating shared understanding through collaborative design", *ACM Transactions on Computer-Human Interaction*, 7(1):84–113.
- Arunachalam, R. and Sadeh, N. 2005. "The supply chain trading agent competition", *Electronic Commerce Research and Applications*, 4:63–81.
- Babb, E. M., Leslie, M., and Van Slyke, M. 1966. "The potential of business-gaming methods in research", *The Journal of Business*, 39(4):465–472.
- Babic, J. and Podobnik, V. 2013. "An analysis of Power TAC 2013 Trial", In *Trading Agent Design and Analysis Workshops*.
- Babic, J., Carvalho, A., Ketter, W., and Podobnik, V. 2015. "Extending parking lots with electricity trading agent functionalities", In Ceppi, S., editor, *Proceedings of the Workshop on Agent-Mediated Electronic Commerce and Trading Agent Design and Analysis (AMEC/TADA 2015)*, pp. 1–14, Istanbul, Turkey. International Foundation for Autonomous Agents and Multiagent Systems.
- Bae, M., Kim, H., Kim, E., Chung, A. Y., Kim, H., and Roh, J. H. 2014. "Toward electricity retail competition: Survey and case study on technical infrastructure for advanced electricity market system", *Applied Energy*, 133:252–273.
- Baetjer, H. 1997. *Software as capital: An economic perspective on software engineering*. IEEE Computer Society Press.
- Bapna, R., Goes, P., Gupta, A., and Jin, Y. 2004. "User heterogeneity and its impact on electronic auction market design: An empirical exploration", *Management Information Systems Quarterly*, pp. 21–43.
- Bell, R. and Koren, Y. 2007. "Lessons from the Netflix prize challenge", *ACM SIGKDD Explorations Newsletter*, 9(2):75–79.
- Bichler, M., Gupta, A., and Ketter, W. 2010. "Designing smart markets", *Information Systems Research*, 21 (4):688–699.

- Chaturvedi, A. R., Dolk, D. R., and Drnevich, P. L. 2011. "Design principles for virtual worlds", *Management Information Systems Quarterly*, 35(3):673–684.
- Cliff, D. and Northrop, L. 2012. "The global financial markets: an ultra-large-scale systems perspective", In *Large-Scale Complex IT Systems. Development, Operation and Management*, pp. 29–70. Springer.
- Coll-Mayor, D., Paget, M., and Lightner, E. 2007. "Future intelligent power grids: Analysis of the vision in the European Union and the United States", *Energy Policy*, 35(4):2453–2465.
- Collins, J. and Ketter, W. July 2014. "Smart grid challenges for electricity retailers", *KI - Künstliche Intelligenz*, 28(3):191–198. ISSN 0933-1875, 1610-1987.
- Collins, J., Ketter, W., and Gini, M. 2009. "Flexible decision control in an autonomous trading agent", *Electronic Commerce Research and Applications*, 8(2):91–105. ISSN 1567-4223.
- Collins, J., Ketter, W., and Sadeh, N. 2010a. "Pushing the Limits of Rational Agents: The Trading Agent Competition for Supply Chain Management," *AI Magazine* (31:2), pp. 63-80.
- Collins, J., Ketter, W., and Gini, M. August 2010. "Flexible decision support in dynamic interorganizational networks", *European Journal of Information Systems*, 19(3):436–448.
- de Weerd, M., Ketter, W., and Collins, J. November 2011. "A theoretical analysis of pricing mechanisms and broker's decisions for real-time balancing in sustainable regional electricity markets", In *Conference on Information Systems and Technology*, pp. 1–17, Charlotte.
- Drew, S. A. 1997. "From knowledge to action: the impact of benchmarking on organizational performance", *Long range planning*, 30(3):427–441.
- EPRI - Electric Power Research Institute, March 2011. "Estimating the costs and benefits of the smart grid - a preliminary estimate of the investment requirements and the resultant benefits of a fully functioning smart grid".
- European Commission, November 2011. "Horizon 2020 - The Framework Programme for Research and Innovation". URL <http://ec.europa.eu/research/horizon2020>.
- Garvin, D. A. 1993. "Building a learning organization", *Harvard Business Review*, pp. 78–91.
- Gilbert, N. and Troitzsch, K. 2005. *Simulation for the social scientist*. McGraw-Hill International.
- Gottwalt, S., Ketter, W., Block, C., Collins, J., and Weinhardt, C. 2011. "Demand side management - a simulation of household behavior under variable prices", *Energy Policy*, 39:8163–8174.
- Gray, D. E. 2004. *Doing research in the real world*. Sage Publications.

- Greenwald, A. and Stone, P. 2001. "Autonomous bidding agents in the Trading Agent Competition", *IEEE Internet Computing*, (2):52–60.
- Gregor, S. and Jones, D. 2007. "The anatomy of a design theory", *Journal of the Association for Information Systems*, 8(5):312–335.
- Hanusch, H. and Pyka, A. 2007. "Principles of neo-Schumpeterian economics", *Cambridge Journal of Economics*, 31(2):275–289.
- Hernandez-Leal, P., Taylor, M. E., Munoz de Cote, E., and Sucar, L. E. 2015. "Bidding in Non-Stationary Energy Markets", In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pp. 1709–1710. International Foundation for Autonomous Agents and Multiagent Systems.
- Hevner, A. and Chatterjee, S. 2010. *Design Research in Information Systems – Theory and Practice*. Springer.
- Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design Science in Information Systems research", *Management Information Systems Quarterly*, 28(1):75–105.
- Hey, A. J., Tansley, S., Tolle, K. M., et al., editors. 2009. *The fourth paradigm: data-intensive scientific discovery*. Microsoft Research.
- Jordan, P. R. and Wellman, M. P. 2010. "Designing an Ad Auctions game for the Trading Agent Competition", In *Agent-Mediated Electronic Commerce. Designing Trading Strategies and Mechanisms for Electronic Markets*, pp. 147–162. Springer.
- Jordan, P. R., Kiekintveld, C., and Wellman, M. P. May 2007. "Empirical game-theoretic analysis of the TAC supply chain game", In *Seventh International Conference on Autonomous Agents and Multi-Agent Systems*, pp. 1188–1195.
- Kahlen, M. and Ketter, W. January 2015. "Aggregating electric cars to sustainable virtual power plants: The value of flexibility in future electricity markets", In *Twenty-Ninth Conference on Artificial Intelligence (AAAI-15)*, pp. 665–671, Austin, Texas.
- Kahlen, M., Valogianni, K., Ketter, W., and van Dalen, J. December 2012. "A profitable business model for electric vehicle fleet owners", In *IEEE Conference on Smart Grid - Technologies, Economics, and Policies*, pp. 1–5, Nuremberg. IEEE.
- Kassakian, J. G. and Schmalensee, R. 2011. "The future of the electric grid: An interdisciplinary mit study", Technical report, Massachusetts Institute of Technology. ISBN 978-0-9828008-6-7.

- Ketter, W. and Symeonidis, A. 2012. "Competitive benchmarking: Lessons learned from the Trading Agent Competition", *AI Magazine*, 33(1).
- Wolfgang Ketter, John Collins, and Maria Gini. 2010. Coordinating Decisions in a Supply-Chain Trading Agent. In *Agent-Mediated Electronic Commerce and Trading Agent Design and Analysis*, Eds: Wolfgang Ketter, Han La Poutre, Norman Sadeh, Onn Shehory, William Walsh, Lecture Notes in Business Information Processing (LNBIP) 44, pages 161--174, Springer Verlag, Berlin, Germany.
- Ketter, W., Collins, J., Reddy, P., and Flath, C. 2011. "The Power Trading Agent Competition", Technical Report ERS-2011-011-LIS, RSM Erasmus University, Rotterdam, The Netherlands. URL <http://ssrn.com/paper=1839139>.
- Ketter, W., Collins, J., Gini, M., Gupta, A., and Schrater, P. 2012. "Real-Time Tactical and Strategic Sales Management for Intelligent Agents Guided by Economic regimes," *Information Systems Research* (23:4), pp. 1263-1283.
- Ketter, W., Collins, J., and Reddy, P. 2013a. "Power TAC: A competitive economic simulation of the smart grid", *Energy Economics*, 39:262–270.
- Ketter, W., Collins, J., Reddy, P., and de Weerdt, M. 2013b. "The 2013 Power Trading Agent Competition", Technical Report ERS-2013-006-LIS, RSM Erasmus University, Rotterdam, The Netherlands. URL <http://ssrn.com/paper=http://ssrn.com/abstract=2268852>.
- Ketter, W., Peters, M., and Collins, J. July 2013c. "Autonomous agents in future energy markets: The 2012 Power Trading Agent Competition", In *Association for the Advancement of Artificial Intelligence (AAAI) Conference Proceedings*, pp. 1298–1304, Bellevue, WA.
- Ketter, W., Peters, M., Collins, J., and Gupta, A. 2016. "A multiagent competitive gaming platform to address societal challenges", *MIS Quarterly*, Forthcoming.
- Kleindorfer, G. B., O'Neill, L., and Ganeshan, R. 1998. "Validation in simulation: Various positions in the philosophy of science", *Management Science*, 44(8):1087–1099.
- Kling, R. 2007. "What is social informatics and why does it matter?", *The Information Society: An International Journal*, 23(4):205–220.
- Koroleva, K., Kahlen, M., Ketter, W., Rook, L., and Lanz, F. 2014. "Tamagocar: Using a simulation app to explore price elasticity of demand for electricity of electric vehicle users", *International Conference on Information Systems (ICIS)*.

- Kuate, R. T., He, M., Chli, M., and Wang, H. H. 2013. "An intelligent broker agent for energy trading: an MDP approach", In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, pp. 234–240. AAAI Press.
- Kuhn, T. S. 1996. *The structure of scientific revolutions*. University of Chicago press, third edition.
- Lang, F., Püschel, T., and Neumann, D. 2009. "Serious gaming for the evaluation of market mechanisms", In *Proceedings of the International Conference on Information Systems (ICIS 2009)*.
- Lee, H. and Whang, S. 1999. "Decentralized multi-echelon supply chains: Incentives and information", *Management Science*, 45(5):633–640.
- Lucas Jr, H. C., Agarwal, R., El Sawy, O. A., and Weber, B. 2013. "Impactful research on transformational information technology: An opportunity to inform new audiences", *Management Information Systems Quarterly*, 37(2):371–382.
- Majchrzak, A. and Markus, L. M. 2013. *Methods for Policy Research: Taking Socially Responsible Action*. Sage Publications.
- March, S. T. and Smith, G. F. 1995. "Design and natural science research on information technology", *Decision Support Systems*, 15(4):251–266.
- Markus, M. L., Majchrzak, A., and Gasser, L. 2002. "A design theory for systems that support emergent knowledge processes", *Management Information Systems Quarterly*, pp. 179–212.
- Melville, N. P. 2010. "Information systems innovation for environmental sustainability", *Management Information Systems Quarterly*, 34(1):1–21.
- Meyer, B. 1998. *Object-oriented software construction*. Prentice Hall New York, second edition.
- Mittelstaedt, R. E. 1992. "Benchmarking: How to learn from best-in-class practices", *National Productivity Review*, 11(3):301–315.
- Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., Kram, T., et al. 2010. "The next generation of scenarios for climate change research and assessment", *Nature*, 463(7282):747–756.
- Nanoha, R. 2013. "Confidence assessment of an agent-based simulation – operational validation of the Power TAC simulation platform", Master's thesis, Rotterdam School of Management.
- North, M. J. and Macal, C. M. 2007. *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation*. Oxford University Press, Inc., New York, NY, USA.

- Palensky, P. and Dietrich, D. 2011. "Demand side management: Demand response, intelligent energy systems, and smart loads", *IEEE Transactions on Industrial Informatics*, 7(3):381–388.
- Peters, M., Ketter, W., Saar-Tsechansky, M., and Collins, J. E. 2013. "A reinforcement learning approach to autonomous decision-making in smart electricity markets", *Machine Learning*, 92:5–39.
- Pries-Heje, J. and Baskerville, R. 2008. "The design theory nexus", *Management Information Systems Quarterly*, 32(4):731–755.
- Pyka, A. and Fagiolo, G. 2007. "Agent-based modelling: a methodology for neo-schumpeterian economics", *Elgar Companion to Neo-Schumpeterian Economics*, 467.
- Reddy, P. and Veloso, M. 2012. "Factored Models for Multiscale Decision Making in Smart Grid Customers", In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*.
- Rittel, H. W. and Webber, M. M. 1973. "Dilemmas in a general theory of planning", *Policy Sciences*, 4(2): 155–169.
- Rosemann, M. and Vessey, I. 2008. "Toward improving the relevance of information systems research to practice: the role of applicability checks", *Management Information Systems Quarterly*, 32(1):1–22.
- Schoder, D., Putzke, J., Metaxas, P. T., Gloor, P. A., and Fischbach K. 2014. "Informations Systems for "Wicked Problems": Research at the Intersection of Social Media and Collective Intelligence", *Business & Information Systems Engineering*, 6(1):3–10.
- Shetty, Y. 1993. "Aiming high: competitive benchmarking for superior performance", *Long Range Planning*, 26(1):39–44.
- Shmueli and Koppius, O. R. 2011. "Predictive Analytics in Information Systems Research", *M I S Quarterly*, 35(3):553–7783. ISSN 0276-7783.
- Simon, H. A. 1996. *The sciences of the artificial*. MIT press.
- Smith, V. L. 1982. "Microeconomic systems as an experimental science", *The American Economic Review*, 72(5):923–955.
- Straub, D. and Ang, S. 2011. "Editors comments: rigor and relevance in is research: redefining the debate and a call for future research", *MIS Quarterly*, 35(1):iii–xi.
- Strbac, G. December 2008. "Demand side management: Benefits and challenges", *Energy Policy*, 36(12): 4419–4426. ISSN 03014215.

- Tesfatsion, L. 2006a. "Agent-based computational economics: A constructive approach to economic theory", *Handbook of computational economics*, 2:831–880.
- Tesfatsion, L. 2006b. "Agent-based Computational Economics: A constructive approach to economic theory", *Handbook of Computational Economics*, 2:831–880.
- Urieli, D. and Stone, P. 2014a. "TacTex13: A champion adaptive Power Trading Agent", In *Proceedings of the Twenty-Eighth Conference on Artificial Intelligence (AAAI14)*, Quebec, QC.
- Urieli, D. and Stone, P. May 2014b. "TacTex13: A champion adaptive Power Trading Agent", In *Proceedings of the 13th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS14)*, Paris, France.
- Valogianni, K., Ketter, W., and Collins, J. 2013. "Smart charging of electric vehicles using reinforcement learning", In *Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence*, pp. 41–48.
- Venable, J. and Baskerville, R. 2012. "Eating our own cooking: Toward a more rigorous design science of research methods", *Electronic Journal of Business Research Methods*, 10:141–153.
- Venable, J., Pries-Heje, J., and Baskerville, R. 2012. "A comprehensive framework for evaluation in design science research", In *Design Science Research in Information Systems. Advances in Theory and Practice*, pp. 423–438. Springer.
- von Hayek, F. A. 1989. "The pretence of knowledge", *The American Economic Review*, 79(6):3–7.
- Wagstaff, K. L. 2012. "Machine learning that matters", In *Proceedings of the 29th International Conference on Machine Learning*, pp. 529–536.
- Walls, J. G., Widmeyer, G. R., and El Sawy, O. A. 1992. "Building an information system design theory for vigilant EIS", *Information Systems Research*, 3(1):36–59.
- Wang, W. and Benbasat, I. 2005. "Trust in and adoption of online recommendation agents", *Journal of the Association for Information Systems*, 6(3):72–101.
- Watson, R. T., Boudreau, M.-C., and Chen, A. J. 2010. "Information systems and environmentally sustainable development: Energy Informatics and new directions for the IS community", *Management Information Systems Quarterly*, 34(1):23–38.
- Wellman, M. P. 2011. *Trading Agents*. Morgan & Claypool.

### **Author Biographies:**



**Wolf Ketter** is Professor of Next Generation Information Systems and chair of the Information Systems section at the Department of Technology and Operations Management at the Rotterdam School of Management of Erasmus University. In addition, he is director of the Learning Agents Research Group at Erasmus (LARGE) and the Erasmus Center for Future Energy Business. Wolf is also the founder and chair of the Erasmus Forum for Future Energy Business. In 2010, he became president of the Association for Trading Agent Research (ATAR). ATAR organizes the annual Trading Agent Competition (TAC). Wolf is leading Power TAC, a new TAC challenge on energy retail markets. He has served as general chair or program chair of more than 20 international conferences and workshops. His research has been published in various top energy, information systems, and computer science journals such as Decision Sciences, Energy Economics, Information Systems Research, Machine Learning, and MIS Quarterly. He serves on the editorial boards of Information Systems Research and MIS Quarterly. In December 2012, he received the prestigious INFORMS Design Science Award, and in June 2013, he received the runner-up award for the best European Information Systems research paper of the year.

**Markus Peters's** research focuses on machine learning algorithms for future retail electricity markets. His work on the topic has appeared in Data and Knowledge Engineering and Machine Learning Journal, and has been presented at various information systems and computer science conferences. Markus obtained his Ph.D. in Information Systems from Erasmus University in 2015. He is now the Director of Development & Operations at Peters Software.

**John Collins** spent 30 years in industry doing research and product development before returning to the University of Minnesota, where he completed his Ph.D. in 2002. There, he taught in the areas of software engineering and artificial intelligence until his retirement in 2013. His research focuses on economic decision processes in autonomous software agents. For the past 12 years, he has been involved in the Association for Trading Agent Research, where he led a major redesign of the supply-chain scenario, served on the board of directors for several years, and is currently managing the continuing development and maintenance of the game scenario and software infrastructure for the Power TAC competition.

**Alok Gupta** is the Associate Dean for Faculty and Research at the Carlson School of Management, University of Minnesota. He is Curtis L. Carlson Schoolwide Chair in Information Management, and the former chair of the Information and Decision Sciences Department. His research has appeared in several

information systems, economics, and computer science journals including Management Science, Information Systems Research, and MIS Quarterly. Alok was awarded the prestigious NSF CAREER Award for his research on dynamic pricing mechanisms on the Internet in 2001, and named an INFORMS Information Systems Society Distinguished Fellow in 2014. He served as a senior editor for Information Systems Research from 2007–2013 and has been serving as an associate editor of Management Science since 2007.