

A framework for identifying suitable cases for using market-based approaches in industrial data acquisition

Torben Jess (tj282@cam.ac.uk)¹, Philip Woodall¹, Duncan McFarlane¹

¹University of Cambridge, Department of Engineering

Abstract

Various researchers have suggested concepts for the application of market-based techniques for data management and data acquisition. However, very few concepts have successfully been implemented and shown a positive practical impact. This can partially be attributed to the lack of existing techniques in identifying appropriate scenarios. Using knowledge from existing successful market implementations in data acquisition, data management and related areas, this paper aims to close this gap. It develops a framework for identifying scenarios applicable for a market-based data acquisition approach. It finds that a heterogeneous environment with distributed decision-making based on partial information combined with assumptions about the users knowledge about the data he is using, can give a good indication for the application of market-based data acquisition approaches. The developed framework is applied towards two sample scenarios.

Keywords: Market-based algorithms; Data Management; Data Acquisition

1 Introduction

In today's Big Data environment, industrial companies are overloaded with data, and it is extremely difficult for decision makers (users) to find the data they need: especially in the case where multiple datasets need to be combined to produce richer data for decision making. These datasets include various internal datasets, or external datasets, which could be bought from an external data provider. All datasets have costs associated with their provision, e.g. maintenance for internal datasets, and acquisition costs for external datasets. Allocating these datasets therefore can be described as an inner company resource allocation problem.

Markets have been shown to work well for resource allocation problems. Various researchers have suggested the application of markets for data management or data acquisition¹ (Christoffel, 2002; Koroni et al., 2009; Yemini et al., 1998). However, very few of these concepts have successfully been adopted in organisations. The main barrier, which hinders this adoption, is being able to identify the right data acquisition and data management problems. Therefore identifying scenarios for the application of market-based approaches can solve this problem and lead to a more successful application of market-based techniques in data acquisition.

¹ This paper focuses on the field of data acquisition in order to reduce the scope of the paper and the analysis. However, various problems in data management can be reduced to problems in data acquisition, making the problem of data acquisition very similar to a list of problems in data management. The selecting of a data quality tool towards a dataset with tool A and tool B as options could for example be reduced to the acquisition of one dataset after treatment with tool A and another dataset after treatment with tool B.

Using interviews to identify industrial cases and literature reviews to identify typical criteria for using market-based approaches, this paper develops a framework to identify promising scenarios. It explains what characteristics need to be fulfilled so that a market-based approach can provide benefits.

A literature review found the following characteristics of market-based approaches: allocation of resources; abstraction of a problem; required fast, efficient, flexible and extensible calculation processes; and incentivising participants for increased efficiency (Brydon, 2006; Koroni et al., 2009; Tucker and Bermann, 1996; Voos and Litz, 2000). This served as input to develop required (definitely needed to apply a market-based approach in data acquisition) and beneficial (not needed but potentially helpful to find scenarios where a market provides benefits) characteristics of data acquisition problems:

Required: multiple data sources; set of alternatives among the datasets; user knows the value of a dataset combination; different data alternatives have costs associated with them

Beneficial: Partial information with data users and/or data providers; Heterogeneous environment for data users and/or data providers; Distributed decision making between data user and data provider

These characteristics can be checked against our scenarios to see if they fit our description for the problem space. For two industrial examples we found good matches using these criteria.

The rest of the paper is structured in the following way. In section 2 an overview about market in industrial data acquisition introduces the research gap, before section 3 shows our developed framework followed by two sample cases in section 4. In section 5 we then present the conclusion and potential future work.

2 Research background

For the research background we are looking into application of markets in other fields and then specifically into data acquisition.

2.1 Industrial application of markets

Adam Smith identified the use of markets in resource allocation and value estimation in his book “the Wealth of Nations” in 1776 (Smith, 2012). Various other researches have shown that markets are good in allocating resources and estimating value (Kaihara and Fujii, 2008; Tucker and Bermann, 1996; Yemini et al., 1998). Markets have been successfully in various industrial resource allocation problems such as supply chain management, workforce allocation (Virginas et al., 2003), airport traffic control (Jonker et al., 2005), or task scheduling (Reeves et al., 2005). They have also been used for intra-company allocation problems; especially helping to solve the NP-complete resource allocation problem (Brydon, 2006). This variety of internal and external applications of markets for companies suggests that they can also provide benefits in other internal areas such as data management (or more specifically data acquisition).

Other application areas related to information systems are robot allocation in manufacturing systems (Dias and Stentz, 2003), bandwidth allocation (Shapiro and Varian, 1998), allocation of CPU and IO capacity (Kwiat, 2002) or in supply chain management systems (Fan et al., 2003).

2.2 Markets in data acquisition

Various authors have developed market-based approaches to data management problems and implementations in related areas. Yemini et al. (Yemini et al., 1998) introduce markets as a concept to application and service resource management for large-scale information systems. But they do not show a concrete application of this market type towards data but rather focus on access to resources and services. Koroni et al. (Koroni et al., 2009) introduce an “internal data market” and the idea that markets can be used for information evaluation, evaluation of information’s quality, costs of information and benefits that information can create. Therefore addressing the key areas within data management. Overall they indicate some potential benefits and challenges but they do not show potential ways to overcome these issues or concrete implementations of “internal data markets”. Christoffel (Christoffel, 2002) describes a market-architecture for data integration, discussing various issues around using markets for data and data integration. However, it has not been implemented or tested, and the specific applicability to test cases is not evaluated.

When reviewing the literature it became apparent that while various approaches for the application of markets towards data acquisitions have been developed, only very few have been implemented such as data security (Dailianas et al., 2000) or allocation of data in a network (Wang et al., 2012) in related areas.

This large number of concepts and the reasonable arguments for applying markets towards specific problems in data acquisition without actual implementation could be due to the complexity of implementing markets or the application towards the wrong type of scenarios where market-based approaches did not show any benefits. The large number of market-based approaches in other fields (see section 2.1) indicates, that the identification of scenarios is the major problem. So far no framework for identifying appropriate scenarios to use markets in data acquisition has been developed.

3 Characteristics of a market-based approach in data acquisition

For the development of our framework we are relying on two main streams. First conduct a review of market-based approaches and the criteria for selecting them. This is then combined with industry knowledge for describing a potential problem space.

3.1 Characteristics of Market-based approaches in data acquisition

Our framework is based on a review of characteristics of various market-based approaches in different fields. We used them to identify the main characteristics potentially applicable to data acquisition. These characteristics can be divided into three categories.

1. The first one are characteristics typical for all techniques calculating the value of a piece of data. (See Table 1)
2. The second one are characteristics often associated and used with a market approach (See Table 2)

3. Characteristics that most likely are going to give a market some form of advantage over existing techniques for data acquisition (See Table 3)

For the first characteristic, the main aspect is about the ability of a market-based approach to get similar results to existing value of information techniques from decision theory and resource allocation.

Number	Characteristic	Description and/or examples
1	Allocation of resources given value and costs (Koroni et al., 2009), (Brydon, 2006), (Xu et al., 2006), (Voos and Litz, 2000)	By combining value of data, users and costs of data providers a market-based approach to data acquisition has the ability to calculate prices between the two sides and make sure the resource data is allocated to the right users in a way that the companies overall utility is increased.

Table 1: Abilities that are similar to existing value of information techniques

Besides these characteristics there are other characteristics typically associated with a market-based approach. These techniques are techniques to formalize users' values and requirements on the one side and supplier's offerings on the other side and then merge them. While these are techniques used in a combination with a market they are not standalone abilities of markets. However, they might increase the benefits of a market-based approach due to the additional calculation required for the automation.

Number	Characteristic	Description and/or examples
2	Formalism and abstraction of problems (Brydon, 2006; Voos and Litz, 2000)	Data managers typically speak of formats, quality, structure of data, etc. of data. Users on the other hand do not always understand this type of discussion. By offering users to buy what they like and offering data managers to provide technical details a market puts the market on a more abstract level.

Table 2: Characteristics of a problem space were techniques typically associated with a market might be beneficial

In addition to these criteria used for a market-based approach a range of additional characteristics of a typical problem space. These characteristics are likely going to be the problem space characteristic due to which a market is able to outperform existing solutions.

Number	Characteristic	Description and/or examples
3	Faster and/or more accurate computation (Brydon, 2006; Tucker and Bermany, 1996;	Meaning a market-based approach for calculating the value of data might outperform other solutions for characteristic 1 of a market-based approach the resource allocation and value calculation.

	Voos and Litz, 2000)	Due to the high complexity it is difficult to calculate the changes in the Value of Information for certain changes in the data. The resource allocation problem is in fact known to be NP-hard. Using a market to transform the resource allocation problem into a winner determination problem (which is still NP hard but has various good heuristics) might help to get better results for the problem.
4	Flexibility and extendibility (Brydon, 2006)	A market-based approach can react better to changes in the data offers and user values. Meaning that a market-based approach could calculate an appropriate and higher overall company utility solution towards the value of information and resource allocation problem under changing conditions in the valuations of users and costs of data offers.
5	Incentivizes participants and increases efficiency (Brydon, 2006; Tucker and Bermany, 1996)	The money (actual or fake) involved in such trading algorithm gives both sides data managers and users incentive to improve their data. It also works as a very honest feedback mechanism towards the data managers and the users. It requires that the data managers orient their development towards the needs of the user and users develop a realistic understanding of their data requirements.

Table 3: Characteristics of the problem space in which a market-based approach might outperform existing solutions.

3.2 Characteristics of the problem space for a market-based approach in data acquisition

Generally a market-mechanism is applicable in a data acquisition context to situations where the value of a piece of data is required in order to make better data acquisition decisions for the allocation of data. Therefore the following conditions should always be satisfied if a market-based approach is considered applicable:

Number	Characteristic	Description and/or example
1	User is using data from multiple (possibly changing) data sources	This characteristics means that the user is using data from more than one data source and that these data sources might change over time.
2	User has a set of offers (from data providers) to acquire more or different data to improve his decisions	This means that the user has offers from data providers where he could acquire additional datasets that would help him in his decision-making. The user needs to select from these offers. This selection will also inform the decision of data provider about the data he needs to provide.

3	User knows the value of a certain piece of data or a combination of data pieces in terms of contribution to a decision	This means that the data is influencing the decision in a way that the user can calculate or estimate its value based on the way the decision changes with this information
4	Data knows the costs associated with its provision	The provision of data costs money and the data managers are aware of these one-off and maintenance costs to provide the data

Table 4: Characteristics of a suitable problem space for a market-based approach to data acquisition

These characteristics describe all situations where a market-mechanism could potentially be deployed. However, in order to provide benefits towards industrial companies a market-based approach also needs to outperform existing approaches. Therefore additional characteristics focussing towards the difficulties that exist within data acquisition and other applications are required.

Number	Characteristic	Description and/or examples
1	Partial information with data users and/or data providers	<p>Data providers don't know about the data interests and data users do not know about the available data and data offers of the other side.</p> <ul style="list-style-type: none"> a. Data providers have no clear idea about how and for what decisions the data is used. They also don't know the value of the different potential data users for a data offer. b. Users have not clear idea about the data that exists or could become available to them. They also don't know the costs associated with providing the data.
2	Heterogeneous environment for data users and/or data providers	<p>Describing situations where:</p> <ul style="list-style-type: none"> a) Users have a range of values for different datasets b) Data comes from different sources and has a variety of characteristics and options to be adjusted c) Users and data managers have a different understanding of problems and approaches
3	Distributed decision making between data user and data provider	Meaning that the decision to provide information to the user is done in a different division (or even company) than the actual business decision (based on data) from the user.

Table 5: Characteristics of use cases where a market can outperform existing solutions

Table 6 shows how these different characteristics match to the initially identified characteristics of a market-based approach, showing an overall good suitability of markets for these identified problem criteria.

Characteristic matching	Allocation of resources given value and costs	Formalism and abstraction of problems	Faster and/or more accurate computation	Flexibility and extendibility	Incentivizes participants and increases efficiency
User is using data from multiple (possibly changing) data sources	✓	✓	✓		
User has a set of offers (from data providers) to acquire more or different data to improve his decisions	✓			✓	
User knows the value of a certain piece of data or a combination of data pieces in terms of contribution to a decision	✓			✓	✓
Data knows the costs associated with its provision	✓			✓	✓
Partial information with data users and/or data providers		✓			✓
Heterogeneous environment for data users and/or data providers		✓	✓		
Distributed decision making between data user and data provider		✓		✓	✓

Table 4: Combination of problem space characteristics with market characteristics

4 Sample cases

In order to validate test our framework we interviewed industrial experts from a large manufacturing organisation in order to gain their expert opinion about potential cases where a market-based approach can be applied. Based on these interviews two hypothetical scenarios were developed. These scenarios are tested against our framework for initial validation.

Scenario A: Company X is a manufacturing company with 100,000 employees. These employees work in a broad range of products and functions. Within this company employee H. Smith in the IT division has the option to acquire dataset containing credit data for suppliers for the costs of 10,000 USD every year. He has a sample data set from the company for one month to test the data set in his organisation. H. Smith thinks that this data could be relevant to some employees in the company, but is uncertain about who specifically could be interested. He knows that the supplier risk analysis division of 10 employees could use this data to predict

potential problems with the supplier, but the quantification of an exact value is difficult. Estimates can be given from the users. H. Smith is unsure whether he should acquire the data set, especially given various other internal datasets, which might help the user to reach a similar decision. Due to the large number of employees, H. Smith cannot ask all potential divisions in the company, due to the large number of employees, distances between employees and due to the fact that he does not know the complete organization.

Scenario B: Company Y is a large machine manufacturing company with 150,000 employees. The service division has 15,000 employees and offers mainly provision of spare parts but also repair services to the companies owning the large manufacturing products of company Y. The customers of Company Y asking for potential spare parts or specific services typically rely on Company Y for support. However, Company Y has little insight into their customer's machines and processes and is therefore unsure about orders of spare parts - orders from customers often "surprise" them. Company Y would benefit from further insight into their customers' data. The problem is that customers would only sell their data for very significant costs of 300,000 USD on average per machine as a customer because the data has high value to them. Manager J. Anderson wants to run a trial activity by acquiring data from 5 customers for 1.5 Million USD. He knows that the supplier forecasting division of 10 employees would benefit from this data, but he is unsure to which extent this data would improve their operation. Each employee in the operation division will be able to give him a rough estimate for the value that this data created for him. However, he is uncertain about the specific divisions and their value for the different part of the dataset. He also does not know where else in the organization this data might be useful.

We checked these scenarios against our initial framework.

Characteristics	Example A	Example B
User is using data from multiple (possibly changing) data sources	<i>The supplier risk division and the procurement division are very likely to use a broad range of data. This data is also very likely to change over time.</i>	<i>Users have different data offers and their data provided from different sources might vary. Especially for the spare parts order in Company Y different data offers could be used to make the predictions. One of them is the customer data.</i>
User has a set of offers (from data providers) to acquire more or different data to improve his decisions	<i>In this case the users have the chance to use the data about credit evaluations for their supplier. Especially the risk analysis department faced with an analytical challenge also has a broad range of data offers they could use for their problems.</i>	<i>Users have various other data offers they could use and in addition have potential customer data. They could use both sources to make forecasts of future required spare parts or just the internal data in case it is cheaper.</i>
User knows the value of a certain piece of data or a combination of data pieces in	<i>Users have an idea of the value data, and its potential combinations with other data sources, would provide.</i>	<i>Each individual user could give estimates when they can use the trial data. This could then be extended for all data sets.</i>

terms of contribution to a decision		
Data knows the costs associated with its provision	<i>H. Smith knows the costs</i>	<i>J. Anderson knows the cost of the data.</i>
Partial information with data users and/or data providers	<i>The users and H. Smith work in are not necessary aware of each other's existence or each other's values and costs.</i>	<i>Data users know the information exists within the customer. However, they do not have access to it and also do not know exactly the value that this data would provide to their respective colleagues and throughout the division, making good estimates difficult.</i>
Heterogeneous environment for data users and/or data providers	<i>Both the data provider H. Smith has a variety of data sources he could look for to provide to the user. In addition the data users have a broad range of potential data offers that can be combined in a different way. Each different combination would offer a different utility to the user.</i>	<i>Each user within Company Y might have a different value for the data, especially when it is combined with other data sources.</i>
Distributed decision making between data user and data provider	<i>The users and the H. Smith work in different departments and make decisions without interacting intensively.</i>	<i>J. Anderson, the customer and the potential data users all work in different divisions or even companies.</i>
Allocation of resources given value and costs	<i>A market can make the calculation between the values given by the users and the costs for the data set known to H. Smith. However this can also be done by other Value of Information solutions.</i>	<i>A market could take the value estimates gathered from the user and combine it with costs of getting the data from each of the customers.</i>
Formalism and abstraction of problems	<i>A market might deploy techniques to automatically capture the users valuation for data and also the users data requirements. This automation could analyze the credit card data and offer it to the users for a trial run. The users would have</i>	<i>Similar to the Example A the users could get a test of the data and then bid for getting this data. The automation could also identify which customer data is interesting for each user to enable the collection of values from the users.</i>

	<i>to give their value in order to keep the data in the future.</i>	
Faster and/or more accurate computation	<i>Due to its ability to deal with the NP-hard resource allocation a market-based approach might outperform existing techniques for the problem of calculating the best allocation of potential data for the supplier risk analysis and procurement challenges to the different users.</i>	<i>Given the 5 data offers of customers data and the large number of users there are various potential combinations of allocating the data to the right users. Calculating which of these combinations generates the highest utility is very difficult, especially because some of the users that might be interested are not know from the beginning.</i>
Flexibility and extendibility	<i>Given an existing market solution of giving the credit analysis data to the users to the resource allocation a market might calculate a new optimum faster once a change to a user value or data providers costs happen (for example another data set for supplier risk analysis becoming available to the supplier risk analysis division)</i>	<i>Once the market did the calculations for the 5 data offer of customer data and the current users he can include additional data offers of customer data potentially building on the knowledge already collected from the existing market.</i>
Incentivizes participants and increases efficiency	<i>Due to the feedback provided from the users for the data offer of having this additional supplier credit information the data provider has an incentive to test additional data offers and overall improve the data provided to all users.</i>	<i>In this case J. Anderson gets the direct feedback form the users whether the data from the customers is worth its price. If it improves efficiency he can continue getting this data if it does not increase efficiency he does not need to buy the data.</i>

Table 7: Characteristics tested against the two example use cases

The initial test show that our framework would cover these two initial scenarios very well. Its current main limitations are that some of the criteria are very general, which can make it difficult to check against them in a scenario. A more detailed checklist breaking down each of the main points could help to reduce this problem. Additional work requires an actual application (beyond initial industry verification) of each of these scenarios in order to verify this framework further.

5 Conclusion and future work

Our paper developed a framework to help identifying suitable scenarios for the application of markets in data acquisition. Addressing the lack of existing techniques in this field. Our initial framework showed promising indications towards the two sample cases presented in section 4. It can help to breakdown a scenario and check if the description of the data acquisition scenarios fits a market-based approach. However, further validation of this framework by

actually applying markets towards the identified scenarios is required. Future work therefore aims to use and refine this framework as part of a questionnaire and scenario description when developing market-based approaches for data acquisition.

In addition to identifying scenarios the framework can also help in the justification process, enabling a clear breakdown for the reasons of applying markets towards a specific data acquisition problem.

References

- Brydon, M., 2006. Economic metaphors for solving intrafirm allocation problems: What does a market buy us? *Decis. Support Syst.* 42, 1657–1672. doi:10.1016/j.dss.2006.02.009
- Christoffel, M., 2002. Information Integration as a Matter of Market Agents, in: *Proceedings of the 4th International Conference on Electronic Commerce Research*. Presented at the International Conference on Electronic Commerce Research (ICECR-5), Montreal, Canada.
- Dailianas, A., Yemini, Y., Florissi, D., Huang, H., 2000. Marketnet: Market-based protection of network systems and services-an application to snmp protection, in: *Proceedings INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies*. Presented at the INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies., Tel Aviv, Israel, pp. 1391–1400. doi:10.1109/INFOCOM.2000.832536
- Dias, M.B., Stentz, A., 2003. *TraderBots: A Market-Based Approach For Resource, Role, and Task Allocation in Multirobot Coordination* (Technical Report No. CMU-RI -TR-03-19). Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA.
- Fan, M., Stallaert, J., Whinston, A.B., 2003. Decentralized mechanism design for supply chain organizations using an auction market. *Inf. Syst. Res.* 14, 1–22. doi:10.1287/isre.14.1.1.14763
- Jonker, G., Meyer, J.-J., Dignum, F., 2005. Towards a market mechanism for airport traffic control, in: *Progress in Artificial Intelligence, Lecture Notes in Computer Science*. Springer, pp. 500–511.
- Kaihara, T., Fujii, S., 2008. A study on modeling methodology of oligopolistic virtual market and its application into resource allocation problem. *Electr. Eng. Jpn.* 164, 77–85. doi:10.1002/eej.20628
- Koroni, A., Redman, T., Gao, J., 2009. Internal Data Markets: The Opportunity and First Steps, in: *COINFO '09 Proceedings of the 2009 Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology*. Presented at the Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology, 2009. COINFO '09., IEEE, Beijing, China, pp. 127–130. doi:10.1109/COINFO.2009.57
- Kwiat, K., 2002. Using markets to engineer resource management for the information grid. *Inf. Syst. Front.* 4, 55–62. doi:10.1023/A:1015386522883
- Reeves, D.M., Wellman, M.P., MacKie-Mason, J.K., Osepayshvili, A., 2005. Exploring bidding strategies for market-based scheduling. *Decis. Support Syst.* 39, 67–85. doi:10.1016/j.dss.2004.08.014
- Shapiro, C., Varian, H.R., 1998. *Information Rules: A Strategic Guide to the Network Economy*, 1ST ed. Harvard Business Review Press.
- Smith, A., 2012. *The Wealth of Nations*. Simon & Brown.

- Tucker, P., Bermany, F., 1996. On Market Mechanisms as a Software Technique (Technical Report No. CS96-513). Department of Computer Science and Engineering, University of California, San Diego, San Diego, California, USA.
- Virginas, B., Voudouris, C., Owusu, G., Anim-Ansah, G., 2003. ARMS Collaborator—intelligent agents using markets to organise resourcing in modern enterprises. *BT Technol. J.* 21, 59–64. doi:10.1007/s10550-007-0082-9
- Voos, H., Litz, L., 2000. Market-based optimal control: a general introduction, in: *Proceedings of the 2000 American Control Conference*. Chicago, IL, USA, pp. 3398–3402. doi:10.1109/ACC.2000.879198
- Wang, T., Lin, Z., Yang, B., Gao, J., Huang, A., Yang, D., Zhang, Q., Tang, S., Niu, J., 2012. MBA: A market-based approach to data allocation and dynamic migration for cloud database. *Sci. China Inf. Sci.* 55, 1935–1948. doi:10.1007/s11432-011-4432-3
- Xu, Y., Scerri, P., Sycara, K., Lewis, M., 2006. Comparing market and token-based coordination, in: *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*. Presented at the AAMAS'06, Fifth international joint conference on Autonomous agents and multiagent systems, ACM, Hakodate, Hokkaido, Japan., pp. 1113–1115. doi:10.1145/1160633.1160834
- Yemini, Y., Dailianas, A., Florissi, D., 1998. MarketNet: A market-based architecture for survivable large-scale information systems, in: *Proceedings of Fourth ISSAT International Conference on Reliability and Quality in Design*.