

From meta data to information architecture

By: [Steven van 't Veld](#), Principal Information Architect

Meta data^[1] is usually seen and treated as an IT issue. In fact it is a business issue. But does business care?

This white paper describes experiences of a practitioner with information architecture and meta data. After a (necessary) introduction, a summary of these experiences is given followed by a description of a practical approach to realize explicit, managed growth of information as a corporate resource.

The main purpose of this white paper is to share knowledge. And to see if we can narrow the existing gap between IT and business.

Information and data

Meta data is usually understood as data-about-data. So: knowing what our data is, and what this data means. When we talk about the meaning of data we get close to something we call information. ISO^[2] defines information as “*any kind of knowledge about things, facts, concepts, etc., of a Universe of Discourse (UoD) that is interpretable to and exchangeable among users*”. In simpler, more practical terms *any knowledge* translates into what we call data. Some of this data we call information in the part of the world we work in (*ISO calls this our Universe of Discourse^[3]: the part of the world we talk about*). Only when we know the part of the world we work in we can differentiate between data that has meaning, and data that has no meaning. So, data with meaning we call information.

Suppose we are discussing the data of our organization (*this organization is our UoD*). We should not be interested in the data that will never be information (*today, or in the relevant past and/or future*) for this organization. It would even be ridiculous to invest in data without meaning because we would spend money on something the investor will never need. So, it is very important for an organization to know which data has meaning. Based on this knowledge it is relatively easy to determine where the organization ought to invest its money.

Based on this observation one can formulate a policy on data and information that is applicable in any organization: “*every organization wants 100% of its information and 0% of all other data*”. Following this line of thinking further one can expect organizations to suffer from a data-overload, but an information-overload must be impossible. An information-overload means a business is not able to digest all its information. Consequently it will have to skip information, and for that reason they are not fully informed. So, an information-overload is in fact a real business risk! One may think “smarter” or “better” IT will solve this problem. This may lead to solutions in some cases, but usually it is not. It is much more likely that we will have to look at the other corporate resources to solve this problem. The organization may, for instance, have to hire more personnel to be able to know all there is to know to do their business.

^[1] In this white paper, commonly used terms are described in different way. This is the result of working in organizations themselves. Most of the concepts are more or less the same, except for the use of the term information architecture. In the absence of a better term we use this term quite different from the way the American IEEE defines it.

^[2] Taken from ISO TR9007 “Information processing systems -- concepts and terminology for the conceptual schema and the information base” and from the final CD for the International Standard ISO/IEC 14481 “Conceptual Schema Modeling Facilities”.

^[3] There has always been a discussion on what Universe of Discourse (UoD) is, what the term means. As a practitioner I have a very practical use for the concept due to the fact I have frequently worked in different environments. A UoD sometimes is a complete organization, sometimes it is a combination of parts of more organizations (for instance in an extended enterprise situation), sometimes it is a part of an organization (terms like domain, department etc. are also used). Having one term to refer to the environment one works in has proven to be very valuable for me as a practitioner. And I have not found a better term, yet, then the ISO term Universe of Discourse (UoD). In this white paper I sometimes use the word environment when I mean UoD for reason of readability. I still think the original term is more precise, and therefore better.

Back to meta data: what really is data-about-data? In line with the above we are not talking about data, but about information. We also do not want data about our information, so it becomes information-about-information. Meta data would in fact be meta information.

In an organization we get used to talk about what we know (*or want to know*) about the information of the organization. We call it the knowledge we have of its information. Is this the same as what we mean by meta data? It should be, but it is usually is not. Knowledge of information is about what an organization wants and needs; meta data is usually about what is available in the IT-infrastructure. Ideally this would be the same, but there still is a very large gap between IT and Business. In fact this is the gap between demand and supply.

Enterprise and Information Architecture

In my experience, knowledge of its information is crucial for the existence, growth etc. of an organization. Based on this observation organizations should treat their information as a corporate resource. An addition to the original corporate resources^[4] Labor (*skill provided by humans*), Nature/Raw Material/Land (*what nature provides*) and Capital (*the goods we created, like machines, transportation equipment, buildings, information technology etc.*). Information has recently become the fourth. Keeping and using optimal quality information in, for instance, communication is crucial for organizations when they want to stay “in business”.

Today, many IT-specialists are talking about IT-Architecture and Enterprise Architecture. They talk about models and structures within their IT-infrastructure, and the value these add to the organization(s) it supports. Recent discussions have put the term Enterprise Architecture into perspective. An Enterprise Architecture is the architecture of a complete IT-infrastructure, including (*system*) management and security, in relation to the organization(s) it supports (*business alignment*). This “definition” is based on many observations of situations where people are talking about enterprise architecture as their profession. One of the consequences of this definition is that enterprise architects are in fact high-level IT-specialists. This conclusion is usually not accepted (*yet*) by the enterprise architects themselves.

The number of approaches to Enterprise Architecture has increased greatly during recent years. A very well-known and accepted approach is the Enterprise Architecture Framework of John Zachman^[5]. This framework relates a list of models of the IT-infrastructure (*mainly in its rows System model and Technology model*) through several Business models to the Scope of it all: organization(s). The Scope row of the framework is closest to what we called Universe of Discourse before.

The problem in the use of these kinds of architecture frameworks in my practice is that it is very difficult to position the knowledge of information as a corporate resource. In the Zachman framework it can only be in the row called Scope. By positioning it as Scope it has become context for the models and structures in the other rows of this framework.

Knowledge of information usually does not fit in the row Business Model. The reason for this is that the models in this row are created during projects to create IT-systems/applications because people working in projects usually hardly ever talk to people working in other projects. So each project creates its own models and structures and it is exceptional when they fit together. These created models are the basis for the meta data. This is the basis why there are problems with meta data; these problems are the inevitable result of the way we work. And there is an even more difficult observation: the knowledge of the information of an organization is usually quite different from the meta data they have for their IT-infrastructure. This is the source of many problems and discussions we have today (*and will have in the near future*).

^[4] Taken from economic theory. In some literature corporate resources are (also) called factors of production.

^[5] For more information, see <http://www.zifa.com/> and other websites.

Suppose we accept information as a fourth corporate resource. The knowledge we have of this information is in fact a specific view on the organization. This view we call the information architecture^[6].

The enterprise architecture contains models, structures and other documentation of the IT of an organization, and its use. This is a view on the IT-solutions that are available. The information architecture contains the knowledge of the information of the organization, and therefore is a view on the business side: from demand. Making our demand and supply explicit enables us to see what needs to be done to create better solutions: a better IT-infrastructure. One of the most important comparisons to be made, here, is between the knowledge of the information and the meta data.

Example: one of the most important parts of the information architecture is out knowledge of the essential entity-types of the organization. So far I have done 20 or 30 of these analyses. We have discovered that organizations always have only between 3 and 12 of these essential entity-types. No more, no less. This has rather large consequences.

Let's look at the organization of a Bank. Say they have between 1.000 and 1.500 business processes. 10.000 applications in their IT-infrastructure is quite normal, so the business alignment is usually very complex. Still, we know a Bank only has 3 kinds of core information (*the essential entity-types*). These 3 have been around for 5.000-odd years, and the fact there are 3 kinds has nothing to do with the way the information is provided to an organization.

Consider this: 10.000 applications basically implement only 3 kinds of (*primary*) information. Try to relate the meta data of those 10.000 applications to our knowledge of these 3 kinds of information. It is not hard to imagine one must be able to simplify the IT-infrastructure based on this knowledge we now have of the information of this Bank in its information architecture.

Creating and developing the information architecture of an organization (*UoD*) usually requires a minimal amount of modeling and structuring. Sure, a conceptual information model would help the demand side greatly in formulating what they really want. Also, some structured way to document the functionality of an organization requires (*maybe implemented as services of the IT*) is important knowledge to have. But the rest of the information architecture is usually unstructured content: information risk analyses, security requirements, policies, planning etc.

An enterprise architecture, in practice, documents the supply side. If we are able to find the "trick" to really relate the information architecture to the enterprise architecture we will be able to compare demand and supply. Meta data will play a very important role to do this. The result will be a great opportunity to diminish the existing gap between IT and business.

Practice

The above may be theory, but practice proves it is very useful to have this theory. I have seen a large number of practical situations where it was impossible to work without.

The following list contains a mixed set of observations that summarize many experiences.

1: Every organization knows its information. A simple, but crucial observation. Organizations do the work that need to be done to realize their targets and strategies. This is impossible when they don't know their information; they would be out of business in no time. A problem is that everybody knows their own part, and usually there is nobody who has an overview. Let alone such an overview is managed. Creating overview, getting consensus on it and putting it under management may be a difficult job, but it is one that adds real value to the organization.

^[6] *The term information architecture is also used by people who create and maintain applications in organizations using web-based technology. This is a different definition than the one used in this white paper. Here information architecture is the view of an organization (better: Universe of Discourse) on its information as a corporate resource. It is the architecture of the information of the organization.*

2: Today the information an organization wants is not (*yet*) what it gets, be it operational information or information for management. Because the enterprise architecture documents the IT-infrastructure it contains in itself not really the right knowledge to change current solutions in solutions the organization really needs. Current (*or near-future*) technology enables current (*or near-future*) solutions. If you want real flexibility, adaptability, agility, effectiveness, usability etc. of the organization and in its supporting IT-infrastructure, start looking at the demand (*what's in the information architecture*) without anticipating very much on existing or expected technology. The Bank example says it can be done. Even stronger: it is impossible to create an optimal, integrated and effective IT-support without really knowing what the organization wants.

3: Let's stop transforming all things-to-be-done in the information- and IT-field into some process, or method. Knowledge of information, for instance, must grow. It doesn't need process oriented management. Of course one can look at and report on the progress of this growth, but you can't introduce "fixed" steps and processes. Managing knowledge is not a project, or even something time-boxed. It's a continuous process, more like what the Japanese know as Kaizen^[7]. Knowledge will have to grow, and managing growth is a continuous task.

4: Outsourcing (*onshore, offshore etc.*) of the knowledge of the information in the information architecture of an organization is not possible. An organization can think about outsourcing IT (*and its enterprise architecture*), but knowledge of corporate resources is something of the organization itself. Outsourcing this knowledge would be "giving away" part of the identity of the organization. Of course an organization can get help in setting up and developing its knowledge, but the organization will need to stay owner of its own knowledge.

5: The information architecture (*knowledge of information*) is the basis for investing in IT development. In practice a specification can be derived by taking relevant content from the information architecture and taking from it what is relevant to know to be able to create requested solutions. An information architecture of higher quality will lead to better specifications. This will diminish the need for an extensive information or business analysis at the start of each IT development project.

6: Investing in new IT will, in future, always lead to explicit procurement activities, independent if the work is to be done by the organization itself or by external contractors (*nearby or far away*). In all cases, a specification taken from the information architecture will be leading. Based on these procurement activities the organization will be able to direct what is to be done. Today, organizations have problems with this direction: the control of investment and exploitation activities. The knowledge of information enables them to control these activities (*again*).

7: Currently there is a lot of discussion on corporate governance and IT-governance. Based on the given separation between demand en supply there is a simple solution: corporate governance manages information demand and IT-governance manages IT-supply. A consequence of this is that IT-management will be positioned at tactical level of the organization. This does not mean IT is not strategic for organizations, on the contrary. But management of capital goods is best done at tactical level.

This introduces a somewhat awkward view on the position of the Chief Information Officer (*CIO*). The name CIO implies it is role at strategic level, a part of corporate governance (*demand side*). But the CIO of today is often an IT-manager. As IT-manager (s)he ought to operate at tactical level, as part of IT-Governance. This touches on an existing problem in current practice where the IT-manager tries to survive at corporate level. It is clear this is a real challenge for most of them.

^[7] Example to read: Masaaki Imai, *Kaizen: The Key To Japan's Competitive Success*, McGraw-Hill/Irwin, ISBN 007554332X

8: As last point: maturity. Practice shows organizations have to be ready to introduce an architecture based approach. This is a maturity issue. It starts when people in an organization begin to realize that they need to do more than invest in IT and exploit the results of these investments. Starting too early usually doesn't deliver good results, supported by the organization. Even stronger, starting too early can easily have negative results, and put further activities on hold for quite a long time.

Starting to introduce architecture with introducing tools, methods and techniques is no real option. In my experience the introduction of tools etc. has backfired many times because the use of these tools tends to become an objective in itself. Without adding real value to what is to be done.

A proven approach

Since the 90's a lot of cases on the introduction of information architecture have become available. They are all different. With a very simple reason: every organization grows in its own areas, and in its own rhythm and pace. Usually there are many reasons why some aspects are developed better, and others are developed less. If, for instance, employees are interested in security, this aspect will be developed better purely because the interest is there, while other aspects are developed less. The reason for these imbalances in growth is usually this simple.

Because every organization develops in its own way, it is not possible to present one-approach-fits-all when you want to develop knowledge on the demand side. In line with the listed experiences there are, however, ways to do things that will let an organization progress.

1. Architecture scan. Based on the use of an extensive checklist in combination with a number of interviews with key people in the organization it is possible to obtain a necessary level of knowledge to go and start work under architecture. Such an analysis usually takes about 15 working days in a 1 or 2 months time period. Together with the result of this scan a number of recommendations on the further development become available for (strategic) management. One of them is usually on how to organize the management of the results of the architecture scan and to enable and promote growth of the knowledge it has "discovered". As soon as this is done an organization usually can start to "work under architecture".

2. Work with the people in the organization. As described before a problem is that everybody usually has her or his own ideas and view about the information of the organization. It doesn't work to go and talk to them, put the findings of these talks on paper and impose them as rules. What does work is to create groups of people from the organization with a broad mix of disciplines/competences. Confront the people in such a group with a large number of relevant aspects, and let them work with them. Together. The experience is they will understand the work to be done better, and they usually open up to each other. It has proven to be a good basis for synergy in the organization. But, as said before, a good mix of different people is crucial in this approach.

With a somewhat remarkable observation. The experience is that IT-specialists are much more likely to have problems with what is happening in these groups than other people involved have. The reason we can think of is that it is very difficult for IT-people to talk about information without talking about tools that will do "things" with this information. This is supported by the view there still is a big gap between IT and business, between supply and demand, between meta data and knowledge of information.

Practical details on this approach

Experience with the described approach can be summed up as follows:

1. An architecture scan usually takes about 15 person days. These days are best spread out over a period of 1 or 2 months. Some extra days may be necessary for presentation, discussion of the results etc.
2. The architecture scan is usually done by one experienced architect. It is a good idea to let him or her work together with one person from the organization itself, preferably the (*proposed*) information architect.
3. The described groups usually contain up to 16 people. Typically it takes 6 days of working together. These days are, ideally planned, 2 weeks apart. The work of the group can, coarsely, be divided in 3 parts: 1/3 of the time is used for introductory presentations on a large number of subjects, 1/3 of the time is spent on discussions about the situation in the organization itself and 1/3 of the time will be spend on the preparation of these discussions.
4. Depending on the organization and their problems the input of specialists may be required in both activities described. Their efforts are, if and when necessary, usually planned on a day-to-day basis.
5. The skills needed in these activities are all high-level. The architect doing the architecture scan needs to be an information specialist. (S)He needs very much, broad knowledge and experience.
6. Which specialists are required depends on the job to be done. Most likely specialists to help out have knowledge in organization culture, business processes and change management.

The outcome of these activities typically is:

1. A full overview of what relevant knowledge is available on business/organization, information and IT.
2. A set of recommendations for strategic management.
3. A proposal on how to organize “working under architecture”.
4. A change in the attitude of people in their work and thinking, thus enabling working under architecture.

The budget for these efforts is strongly dependent on the number of groups that need to be formed. Basically we are talking about an investment of between € .75K to € .100K for both activities. The cost becomes higher when more groups are formed.

The time necessary, again this depends on the number of groups to be formed, is typically somewhere between 6 and 9 months. This time depends heavily on the speed an organization can develop with respect to the work. Going too fast usually creates big problems.

Concluding remarks

Meta data is usually seen as something of and for IT-specialists. The reason is that most problems with meta data are strongly related to what is happening in and around the IT-infrastructure. Problems with knowing which data is where, bringing data together for all kind of purposes (*operational, managerial etc.*) and many other issues are problems that are related to meta data.

In this white paper, meta data is related to the knowledge of the information of an organization. This knowledge is the basis for the information architecture of the organization. In fact, the information architecture ought to be the basis for the meta data in the same way as the demand for information is the basis for the supply of solutions. Ideally, differences between demand

and supply should not occur. When they do occur it means supply does not meet demand (*yet*). Or: the IT-infrastructure does not do (*yet*) what the organization wants it to do.

By developing and managing the information architecture of an organization the demand for information of this organization can be made explicit. Having good knowledge of what the demand is will be leading for the definition of meta data itself. Really knowing what the demand is, is crucial for a real integration of the components of the IT-infrastructure and for an optimal support of a flexible, agile organization.

For these reasons, and many more, the information architecture is crucial for an organization to obtain real control and management over its IT-infrastructure. The problem is that most organizations still think and work in projects on IT. Experience is that the described information architecture will put all of this in the normal perspectives an organization has and works with.

Steven van 't Veld

Steven is an independent, highly experienced, skilled and qualified Principal (*Enterprise*) Information Architect with experience of operating within a wide variety of markets across public, private and SME sectors. He is a recognized leader in his field and has a proven ability to consult and advise at all levels and within all types of organization. He is skilled in coaching and mentoring employees at all levels to improve performance and he is experienced in (enterprise) information architecture, conceptual information modeling, document management, procurement of very large IT-systems, knowledge management, data-warehousing, organizational scanning, due diligence, presentation, team leadership and management, quality, efficiency, productivity and cost-control initiatives. He has worked in ISO standardization activities on Open Distributed Processing, Application Portability and Conceptual Schema worldwide between 1985 and 2000. He co-ordinates Pan-European activities on forming of professional bodies for architects and chairs an independent organization for the certification activities for IT, Enterprise and Information Architects.

A/I/M bv is an independent management consultancy firm based in the Netherlands, specialized in information architecture. The services A/I/M bv provides are consultancy, coaching and Master class-level courses on Information Architecture, Conceptual Information Modeling etc.

Contact through E-mail: Steven.van.t.Veld@AIM.nl

More information on: <http://www.aim.nl/>