141. Commit-Time Requirements for an Ontology Server

Mohammad Nazir Ahmad
School of IT & Electrical Engineering
The University of Queensland
nazir@itee.uq.edu.au

Robert M. Colomb
School of IT & Electrical Engineering
The University of Queensland
colomb@itee.uq.edu.au

Abstract
An ontology is recognized as the solution for the integration of information systems. The environment of interoperation may involve many players who have agreed to commit to the ontology in order to maintain their system of speech acts and institutional facts in conformance with the coordinated system. When this interoperating community is established, it can generate a large number of institutional facts due to different range of players who can request different facets of information. In this light, however, how those players commit to ontology is still unclear. This paper discusses what sort of requirements that we need to assist how the players commit to the ontology. The approach of this paper is theoretical which is based on the literature of the concepts of speech acts and institutional facts and a case study of the Olympic games. As a result, we have defined several important commit-time requirements to explain situations of players committing to ontology in the context of ontology-based interoperaction of information systems.

Keywords: Systems Integration, Commit-Time Requirements, Ontology, Ontology Server, and Interoperability

Introduction
Sheth and Larson (1990) systematized the research into Federated Databases. However, its failures to systems integration generally foundered on the problem of semantic heterogeneity – that similar terms in different systems generally have related but different meaning (Colomb and Orlowska 1995; Colomb 1997). The failure is explained by the relatively recent philosophical theory of institutional facts (Searle 1995). Further, Colomb (2004) proposes the necessity for business process reengineering to achieve integration of information systems within organization, and prediction that inter-organizational integration of information systems using the Internet can succeed only if the applications share their systems of institutional fact. One of the major uses of ontology is in fact to support interoperation of information systems (Hart et al. 2004). Once an interoperating community is established can generate a large number of institutional facts. These institutional facts can be interpreted by any of players (i.e. human or software agents) who share the common ontology. In this light, those players committing to ontology basically agree to maintain their system of speech acts. However, how these players commit to this ontology needs to be identified. Therefore, this paper defines significant commit-time requirements to support ontology-based interoperation information systems. In a concrete point of view, we have an ontology and throughout the paper, the concept of players refers to a user responsible for an application that is to commit to the ontology. The ontology consists of a collection of objects (i.e. classes, individuals, properties, property instances). The user employs the browser to identify the objects the user considers needed to commit the application. Nevertheless, this study is only concerned to discuss from an institution’s point of view (organizational and business) and uses the Olympic Games as a case study to support our analysis and some example given is basically analogous to the concept of players and situations when committing to the ontology.
A Background and Theoretical Framework
This section provides useful background information and considered as our theoretical framework of our study. It is important to understand some fundamental concepts like ontologies, conceptualization and specification, systems of institutional facts, institutional world and interlocking institutional worlds.

Conceptualization and Specification
Gruber (1993) defines “An ontology is a specification of a conceptualization”. Conceptualizations are abstract entities that only exist in the mind of the users or communities. In order to be documented, communicated and analyzed, these entities must be captured in terms of concrete artifact. The representation of a conceptualization is called a specification – the ontology. Note that the problems of semantic heterogeneity, it is largely arises from the conceptualization, not the specification. The conceptualization is independent of the languages unlike the specification. There are many representation systems like OWL, UML, ER and Topic Maps used to represent conceptualization explicitly. In fact, many Computer Aided Software Engineering (CASE) tools developed deal with specification and not conceptualization.

The Speech Act and Institutional Facts
Searle (1995) proposes there are two types of facts; brute facts and institutional facts. A brute fact is a fact that exists independently of any human society or even human cognition, for example the facts of the physical sciences. Thaksin became ex-Prime Minister when the King uttered certain words in a particular context. The utterance of certain words is a brute fact. Searle (1995) points out that although we live in a world of physical objects, many of these objects have significance dependent on human society. Another example, an Australian $1 coin is a metal disk. In Australia, many people are willing to exchange one of those disks for a candy bar. In the US, or Europe, or China, if someone were to offer to exchange the same disk for the same candy bar, the best that could be expected is a strange look. The disk is money in Australia, but not in the US, Europe or China. The metal disk as a physical object is called a brute fact. That the disk is money in Australia is called an institutional fact. The two tied together by context. The institutional fact came into being when the brute fact occurred in a particular institutional context. Searle (1995) uses a formula: “(brute fact) X counts as (institutional fact) Y in context C”. The world is full of institutional facts. A student’s name enrolled in particular course is an institutional fact. The brute fact is a course registration form. Entering a student’s name is an act, called speech act, probably performed by student and a registrar acting in cooperation during the course registration process.

An Institutional World and Interlocking Institutional Worlds
Each institution creates a collection of institutional facts (instances of institutional fact type) created in speech acts. This is called an institutional world. In Olympics, the International Olympic Committee (IOC), National Olympic Committee (NOCs), Organizing Committee for a particular Games (OCOG), International Sports Federation (ISFs), are institutions, autonomous though loosely-coupled. Their institutional world is the conceptualization of which the ontology is the specification. When two organizations share their systems of institutional facts, that two institutional worlds interlock, therefore they can interoperate. Two institutional facts interlock in several ways. The following list is not intended to be exhaustive: 1) the speech act creating them is performed by both institutions, i.e. a purchase involves a buying speech act by one institution and a selling speech act by the other; 2) an institutional fact in one institution is part of the context of a speech act performed by the
other, e.g. a student's speech act of enrolling in a course depends on the prior speech act of creation of the course by the University. The institutional fact of the course is part of the context of the student's enrolment; 3) A speech act by one institution can be performed under licence by another. A taxi ride is an example. So is the conduct of sporting events by an OCOG (the rules of the events are specified by the sporting federations, so the OCOG is in effect acting as an agent of the sporting federations); 4) One institution may constrain the speech acts of another, e.g. the IOC constrains all events in the Olympics to award gold, silver and bronze medals, a constraint on the rules of the event created by the sporting federations. The Olympics is a good example of interlocking institutional world because they come together every four years in a different configuration and interlock to make a games success. In this light, an organization’s committing to the ontology is exactly agreeing to maintain their system of speech acts and institutional facts in conformance with the coordinated system. For further clarification, the concept of interlocking institutional worlds is elaborated deeply in (Colomb and Ahmad 2007).

**Defining a Commit-Time Scenario**

We have an ontology and a user responsible for an application that is to commit to the ontology (i.e. Olympics ontology). These users are probably running for many ranges of institutions and basically more than it appears from our Olympic scenario. This world (Olympics) is actually very much larger. The Organizing Committee for a particular Games (OCOG) has to interact with the main institution in the Olympics (i.e. IOC), negotiating with the ISF(s) and communicating with NOCs. The OCOG has also a large number of functional groups that need to interact with a large number of external bodies to make the games a success. These external bodies range from television broadcasters through the security services to the caterers for the Olympic village. We can assume a large number of users or players here committing to ontology as shown in Figure 1.

![Figure 1: Committing to Ontology](image)

Figure 1 shows that every player may have their own conceptualization (C) since differences in their own way of doing things (i.e. P₀>C₀, P₁>C₁). Furthermore, each fragment (F) of the ontology relates to the respective player who contributes to the entire ontology (i.e. P₀-
>F₀, P₁>F₁). On one hand, the respective player can create its various institutional facts (i.e. IOC determine what an Olympic medal), and other players can see them. On the other hand, this creation of institutional facts can be changed. Any changes might need to be informed to other players by means of announcement or queries. Moreover, players may have their own way to see the ontology. They would integrate some information from different fragments of the whole ontology. They would also have their relevant information from the ontology. The overall ontology (Olympic ontology) is actually the result of interlocking institutional worlds, therefore interlocking ontologies. In order to commit to the ontology those players must adjust their way of doing things conform to the Olympics. Colomb (2004) claims that this requires much negotiation and resolution among more or less equal partners. In some cases, one of the institutions will have greatly more power or prestige than most of the others, so will be able to more or less dictate the resolution. Before joining, the semantic heterogeneity (i.e. arise from each conceptualization) must be resolved each time by negotiations amongst players. This section extends our analysis of commit-time requirements as argued in (Ahmad and Wahid 2006b). It has lead to our understanding of commit-time situations. The next section draws an extended example of Olympic scenario to illustrate the commit-time requirements.

A Commit-Time Requirement - the Olympics Example

Even though institutions are generally more or less autonomous, they do sometimes cooperate with each other. Again, we will use the Olympics scenario as examples. The central institution in the Olympics is the IOC. The IOC authorizes the creation of NOCs in countries and territories. The IOC is in turn constituted partly by the NOCs. One of the tasks of the IOC is to select an Organizing Committee for a particular Games (OCOG). The selection is made from a set of candidate OCOGs each selected by their NOC from bid committees from possibly several competing cities from that country. For example, the Chinese Olympic Committee (COC) bid was successful and has led to host the Games in Beijing 2008, resulting in the constitution of the Beijing Olympic Organizing Committee (BOCOG). The COC is confronting two responsibilities - closely co-operating with BOCOG to ensure a successful Games and achieving overall improvement in China's sporting scene to better Olympic results in 2008. It has to enforce the requirements and working standards set out by the IOC and individual ISF to create a regulated and just competition environment. The BOCOG at the time of writing are still negotiating with relevant international sporting organizations to determine the facilities for tennis and beach volleyball. Returning to the example at hand, we aim to define four important requirements; level of granularity, relevant portion, separation of concern, and publish/subscription of the portion.

Requirement R1 – the Level of Granularity of the Portion

It is common to a respective player to access a different level of details of the institutional fact. From database point of view, this player may have a CRUD (create, read, update, delete) operation for the different granularity of information structure. Take an example of badminton court. The respective institution like IBF has an exclusive right in term of CRUD operation to change the court at different level of granularity. It can update the classes, rules, and properties of the court, and in fact into the level of individual. Besides that, a respective player can also view a different level of details of the portion. For example, the IBF might view the court range from broad (think of the IBF might want to provide a summary of the court specification in order to explain how badminton player can score the points) to detail in the specification of the court (think of the IBF might want to provide a
details of the court specification in order to supply information to the OCOG who want to award a contract to the contractor). Another player like a contractor who builds this court supposes to know in detail the specification of the court.

![Diagram of BadmintonCourt](image)

**Figure 2: An Institutional Fact, BadmintonCourt**

**Requirement R2 – the Relevance of the Portion**

It is common to players to have relevant information of the institutional fact. It is costly in terms in resources to understand the huge information that most of them are irrelevant. Therefore, the ontology does have to be relevant portion to the players and which are not relevant to that player should be discarded. Bhatt et al (2004) say that using the whole ontology means that drawbacks from this redundant information are encountered; complexity and redundancy issues rise, while efficiency issues fall. In a real world, physical objects are dependent on other objects. For example, humans are dependent on food. Information and especially descriptive information can also have dependency. The dependencies in information refer to “what you need to know about in addition to what you want to know about”. In a more practical example, if you want to troubleshoot why your printer does not work, you would need ontological information about that printer. Of course you probably might have its manual procedure or you can search on the Internet. However, the information about the printer system requires that you understand what type, series or model of your printer you have.

In Olympics, different user groups such as athletes, coaching teams, organizers, spectators and other groups of people from all over the world need a broad range of information; sports, transports, logistics and tourism are only the most obvious application domains. Furthermore, all these will be provided by a large number of suppliers, ranging from the IOC, NOCs, ISFs, BOCOG over public authorities to private companies in China as well as from abroad. Therefore, it is crucial to provide relevant portion of the entire ontology to these players without overloading them with irrelevant portions. Take of spectators as examples. The Olympics is based on individual competition and consists of a single-elimination tournament. Each match is played to the best of three games. The spectator might interest in Olympics badminton competition. The ontology may store information about BadmintonCompetition. It may have four types of competition, Rounds enable each competitor plays in at most one fixture per round. As rounds progress the lost competitor is eliminated and the last 16 individual competitors for each event is listed. These 16 competitors will be eliminated to 8 successful competitors in
QuarterFinal. In SemiFinal there are 4 competitors will compete to get into the Final. It is obvious to particular spectator in order to know the badminton results and competition being held, for example. There are some other related information that he might need to know like the entire badminton competition (subtype structure) and which other classes that related (dependency relationship) to badminton competition including which type of event, schedule, venue and in fact transport information.

![Figure 3: Olympics Ontology Example](image)

**Requirement R3 – the Separation of Concern of the Portion**

The concept of Separation of Concern (SoC) is quite common in computer science and was probably coined by Dijkstra (1982). Colomb (2004) states where the interoperating community consists of many players, there may be an advantage to each player giving its exclusive access to the institutional facts it creates in favour of a community-wide pool to which all players have common access. He added that this is common for example in real estate, where individual sales reports and auction success rates can be published for a whole city market are, enabling each player to see trends, to which they can respond in their own fashion. To our mind correctly, enable players to have their own immanent ontology using the common transcendent ontology (Olympics) is one of the commit-time requirements. Due to different players might require different organization of information and different vocabularies to suit their information seeking needs. We define here that the portion which players access can be reorganized by that players according to their need. Reorganization of that portion is actually local to the player’s context. However, any changes in the global portion should propagate the player’s view. Analogously, we can think of designing an individual professor’s site. The information structure for each site is almost same (similar ontology). However, we might have different style of sites in term of layout and interaction. Each web designer may have different classification system or ways to organize the same information. The ontology of news topics in a newsfeed change as events happens in the world. The ontology structured in the directory of a person’s laptop is immanent, because the user of the computer is free to change the directory structure. The organization of information here is called, views. Note that the views not only relate to the same player’s portion but can be overlapped to more or less portion of the ontology (think of OCOGs integrate information range from sports dimension to venue and Olympics village). For the same portion, consider an example of Olympic badminton result. An ontology could be in a portion as shown in Figure 4:
The media broadcasters play an important role in Olympics. They may range from the members of accredited written and photographic press and Non-rights-holding Broadcasters (ENR) as well as coordination of broadcasting services for Rights-holding Broadcasters to Host Broadcaster-Beijing Olympic Broadcasting Co., Ltd (BOB). The information about results might be presented in many ways to the users. It could be presented starlist by NOCs, participation by NOCs, medallists, and detail events and so on. The followings simple examples of “Separation of Concern -SoC” perceived by two different players.

**Requirement R4 – the Publish/Subscriotion of the Portion**

It is clearly understood that when we have a huge ontology, many players require a portion of the ontology. The relevant portion consists of interrelated classes that relevant to a given player. However, it is also common those classes subject to be accessed by more or less players. Many cases there are some classes are visible or public to a player but invisible or private to other players. The notion of public and private is very familiar in software engineering particularly in object-oriented programming. We refine the public to refer visible classes and the private to the invisible classes purposely in the context of player’s point of view. We would adopt these notions to facilitate the need for publish/subscription of the portion by any of many players. The ontology fragments the participant accesses will be
subject to change. For example, when a new version of the ontology is published, we will need to be able to decide which participants need to be informed of the changes. One reason a participant might need to know about a change is because the fragment of the ontology to which the participant subscribes changes. A second reason might be that a player might wish to be kept informed of changes in nominated areas of the ontology even though they do not at present make use of these aspects in their interoperations. Notice that the institutional facts in the various conceptualizations are actually created by their respective institutions in speech acts. Only one body determines what an Olympic gold medal means or how a badminton event works. The participants in the interlocking institutional worlds see the institutional facts created by the other participants as fixed reality. These institutional facts are created by what are called performative speech acts. Because they are human creations, they can be changed. When an element of an institutional world changes, the other participants need to know about it, which they do by either an announcement being made or by means of a query. Both announcements and queries are also speech acts, called informative. The decisions made which allow the various institutional worlds to interlock are also performative speech acts. The interlocking institutional world is created by performative speech acts and held together by informative speech acts. The performative speech acts are reflected in the specification by the creation of classes, properties, individuals and statements. The informative speech acts are reflected in queries and notifications of change if the server software supports a publish/subscribe facility.

Take a simple example of scoring system in badminton. In May 2006, the sport's governing body, the IBF, changed its institutional world by changing the scoring system from service-based points (3x15) to rally-based points (3 x 21) to make the game faster and more entertaining. The main change from the traditional system was in which the winner of a rally scores a point regardless of who served; games were lengthened to 21 points. This new “statement” about changes should be propagated to all players that subscribe to this portion mainly the BadmintonEvents. In summary, this new system defines; Games: Each match is the best of three games; Points: A game is won by the first side to score 21 points; Exceptions: If the score reaches 20-20, the winner is the player or team with a two-point advantage. And if the score goes up to 29-29, the winner is the first to reach 30 points; Rally not Serving: a player or team can win the point without holding serve. Many players such as referee, badminton players, NOCs and BOCOGs should be announced/may query about the changes of institutional fact (Badminton Events and Rules).

![Figure 6: An Example of Publish/Subscription of the Portion](image)

Rule1: The winner of a rally scores a point regardless of who served.

IBF changes rules (publish)
A Necessity for Mediation Services – the CASE tools requirements

It is well understood that the “business requirements” that we have defined in section 3 and followed by section 4 would be implemented in term of “software requirements” for the ontology server. This server is closely similar to the development of relevant CASE tool for designing information systems. It can be clearly seen that the necessity for mediation services as the server facilities to accommodate more or less players commit to the ontology. These facilities might range from functions like – browsing, selecting, subscribing, publishing, translation services, extracting and communication protocol of the portion of the ontology. The operations would be focused on part-centered approach to the ontology. The discussion of these functional facilities is out of the scope of this paper. However, we discuss briefly one of them that show what is needed for communication facilities.

Consider an example of Olympics medal. In Olympics, competitions have medals awarded to the top three competitors: gold, silver and bronze. Race-oriented disciplines like swimming and track have an easy way of identifying the first, second and third to finish. So do performance-oriented disciplines like pole vaulting, weight-lifting and trapshooting. And so do judgment-oriented disciplines like diving, gymnastics and synchronized swimming. But some disciplines are match-oriented, like water polo, soccer, tennis, badminton and boxing. The normal organization of competitions in these sports is as tournaments, where the winner is the competitor who wins the final match. There is no natural way to determine a third-place winner. Even some events within normally race-oriented disciplines are match-oriented. Think of pursuit cycling. The common approach is for the event to be organized as a tournament with a pair of semifinals, the losers of which play a match for the bronze medal. This requires the sporting federation to change the event’s rules, therefore change the event’s ontology, in order to successfully merge with the Olympics ontology.

Figure 7: An Example of the Mediation of the Changed Portion
This also applies to other players like television broadcasters. For example, Beijing 2008 will be broadcasted worldwide by a number of television broadcasters. Confirmed broadcasters include: Mainland Chinese state-owned CCTV, predominantly CCTV-5 will have exclusive coverage rights, China will also be streaming all the events over the internet showcasing the China Next Generation Internet, Channel Seven in Australia, CBC and Radio-Canada and its properties, along with TSN and RDS, in Canada, NBC Universal, with NBC and its cable properties, in the United States and BBC in the United Kingdom. Some athletes, especially swimmers, have voiced dissatisfaction with the IOC's decision to schedule some events to meet the requests of NBC, which paid US $3.5 billion for exclusive United States broadcasting rights to the summer and Winter Games from 2000 through 2008. NBC requested that popular events, such as swimming, athletics, basketball, and gymnastics, be broadcast live and during television primetime in the U.S. (i.e. 6 to 9 p.m.) for maximum advertising revenue. This would require events to be held in the early morning, Beijing time. The IOC granted the request concerning swimming and gymnastics but denied it for athletics and basketball.

The simple examples given illustrate the necessary to mediate player to join the ontology. The mediation process is involved negotiations amongst more or less power players in order to dictate the resolution. In the case of medal, the IOC is more power than sporting federations. In practical, it would be useful to define communication rule in order to mediate the process of committing specifications of conceptualization to the overall ontology.

Conclusions

From the perspective of institution, the Olympics is a good example of an institutional world formed by the interlocking of many institutional worlds because it is large and complex, well established, familiar and very well published, especially in its sporting dimension, but also in its business dimensions due to the requirement for accountability to all the stakeholders. This larger institutional world is the conceptualization of which the ontology is the specification. This paper discusses how more or less players commit to this ontology. The player here analogous to a user who is responsible for an application that is to commit to the ontology. We conclude that to accomplish ontology-based interoperation of information systems require at least four functional commit-time requirements. To fulfill these requirements, there are at least three things necessary: (a) the data structures must be rich enough to represent the complex semantics; (b) there must be flexible customization mechanisms that enable multiple players to view the portion of the entire ontology; and (c) there supposed must be mediation facilities that assist the players to “see the world – the ontology”. We believe that ontology is the answer to the (a), ontology views are the key to the (b) and ontology server is a main tool to the (c). For the future works, we will devise some technical solutions for the various requirements. We plan to propose our design framework to fulfill these requirements. We will develop metamodel for ontology server based on our ontological analysis from relevant literatures on formal ontology argued in (Guizzardi, 2005; Guarino et al. 2004; Wand et al. 1990). We will apply the MDA to ontology approach (software engineering paradigm) to the development of this metamodel using UML MOF and UML ODM profile as proposed in Colomb et al. (2006). The server is a kind of information systems which are closely related to the research in CASE tools. If the CASE tools are generally used to support the design of system, on the contrary, the ontology server is used both during design, commit and in fact during the execution of the ontology-based applications as reported in (Ahmad et al. 2006a; Ahmad and Colomb 2007)
References