



Netzbasierte Informationssysteme **Semantische Middleware** **Topicmaps**

Prof. Dr.-Ing. Robert Tolksdorf
Freie Universität Berlin
Institut für Informatik
Netzbasierte Informationssysteme
mailto: tolk@inf.fu-berlin.de
<http://www.robert-tolksdorf.de>



Towards a tuplespace-based middleware for the Semantic Web

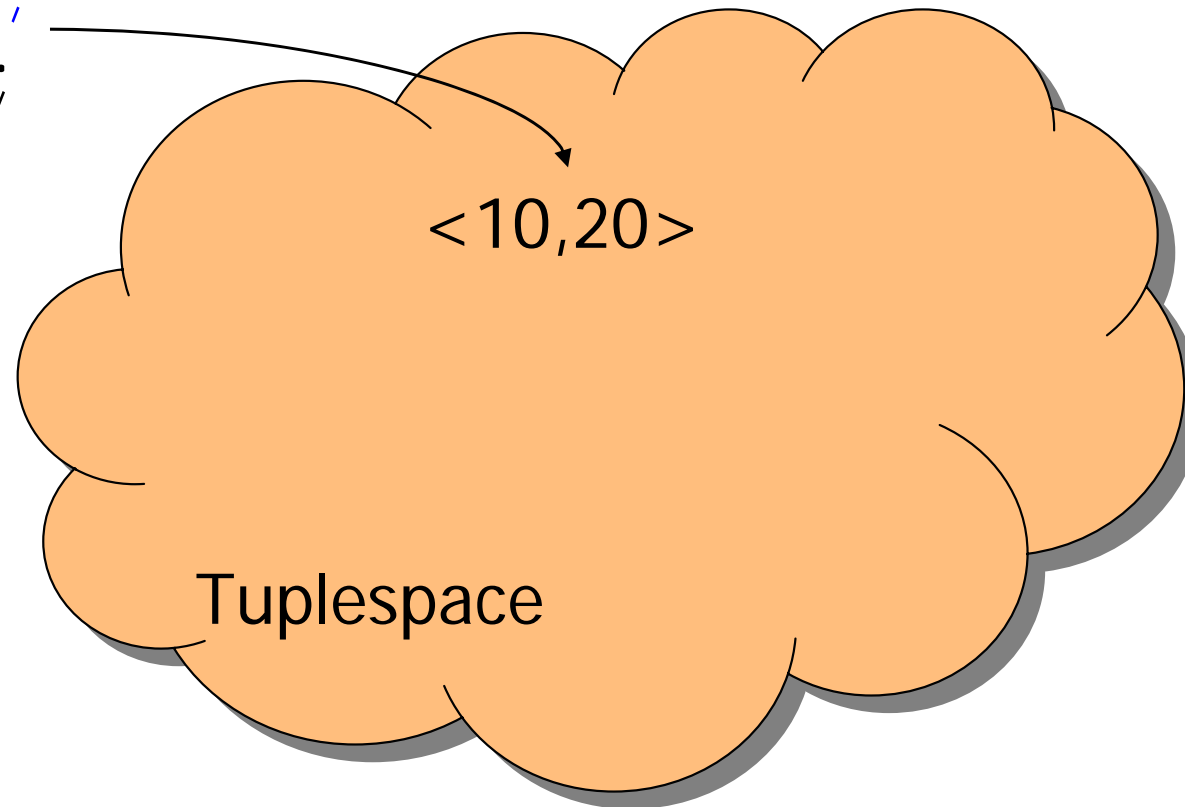
*Robert Tolksdorf, Elena Paslaru Bontas,
Lyndon J. B. Nixon*

- In the Semantic Web playground we find
 - basic Web technologies like URIs, HTTP, XML...
 - models and languages to formally express them like RDF, OWL...
 - lots of components like Sesame, existing ontologies,...
- We miss a glue that
 - connects the components (and the data) together
 - offers a high level abstraction to coordinate them
 - is open and can incorporate heterogeneous components
- Middleware for the Semantic Web is needed that is
 - decentralized, not centralized
 - scalable, not limited
 - high-level, not technical

- Coordination language Linda [Gelernter and Carriero, 1992]
 - tuplespace, some simple primitives
 - Origin: Parallel computing, also used for
 - Distributed systems
 - Service oriented open systems
 - Web based systems
 - Workflow systems
- Variety of extensions to Linda
 - 20 years of research
 - Linda and
 - FORTRAN, Prolog, C, Objective-C,...
 - persistency, databases, transactions
 - agents, objects, scripts, rules
 - ...
- Of special interest for Semantic Web
 - Linda and logic programming
 - Linda and rules
 - Linda and scalability

Linda: Indirect interaction

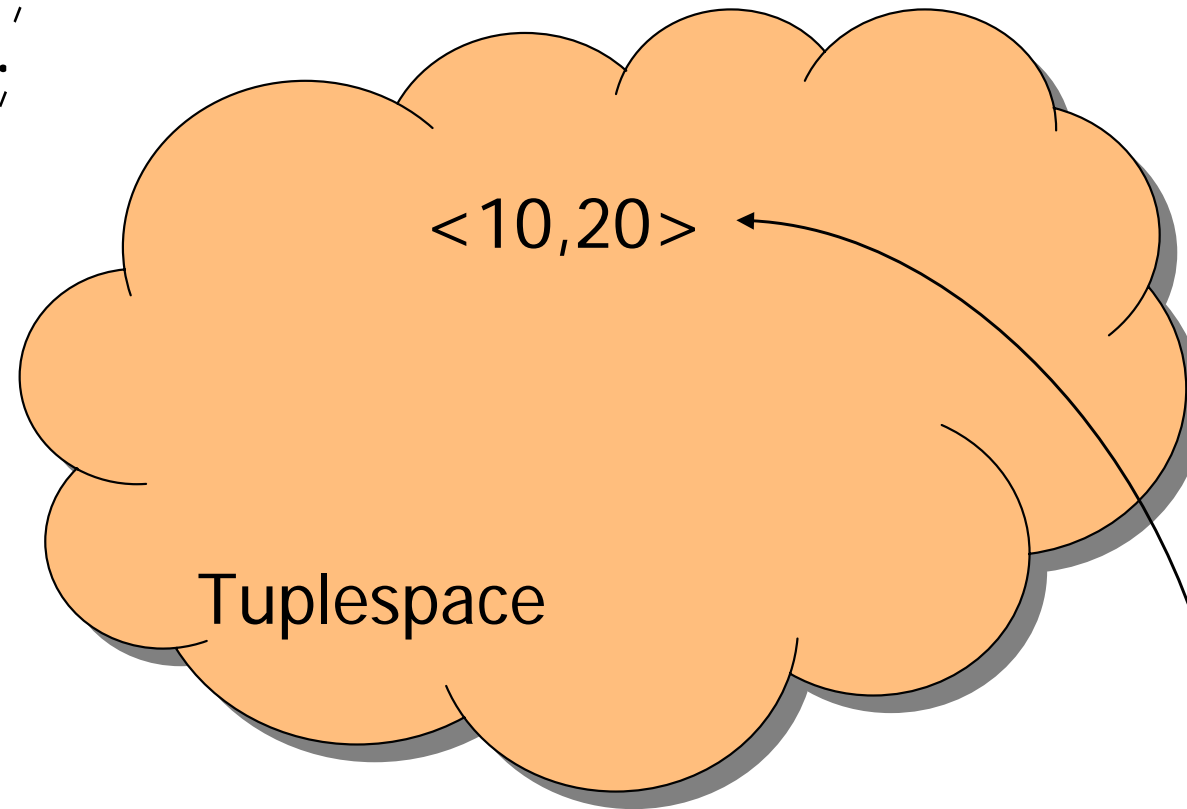
```
out(10,20);  
in(?result);
```



```
in(?a,?b);  
out(a+b);
```

Linda: Indirect interaction

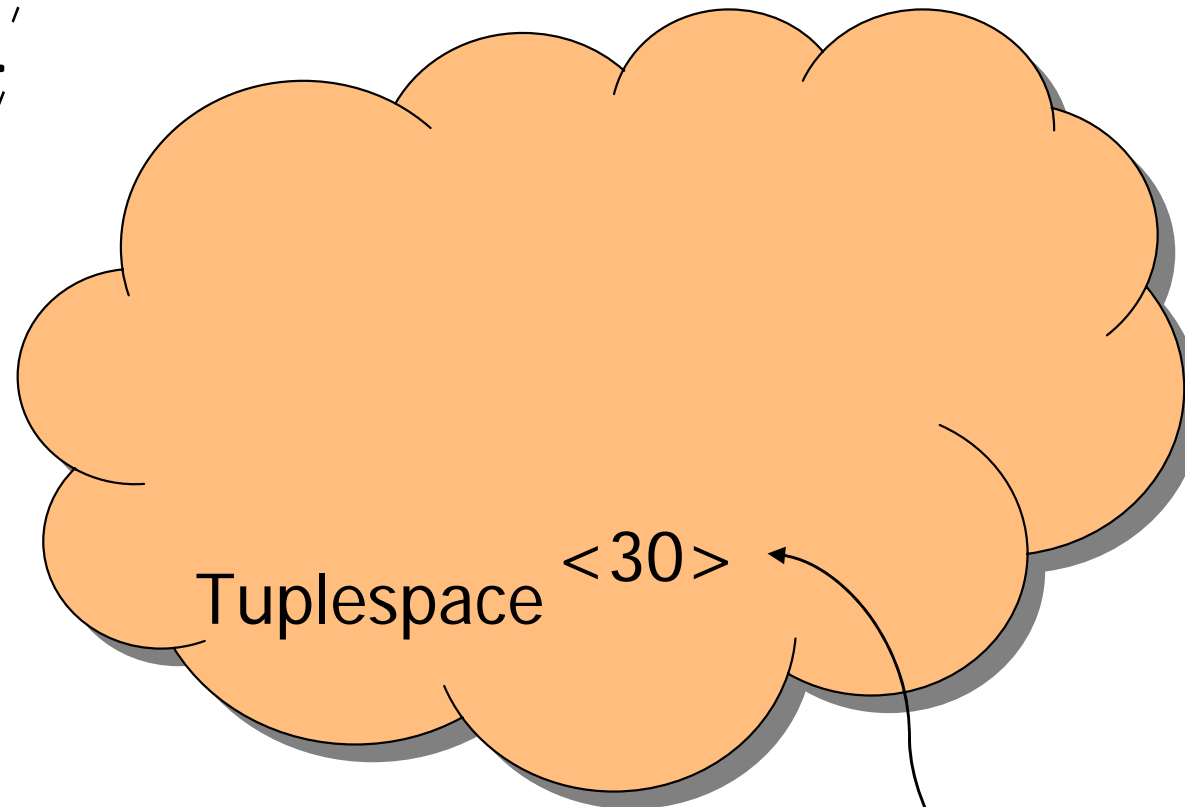
```
out(10,20);  
in(?result);
```



```
in(?a,?b);  
out(a+b);
```

Linda: Indirect interaction

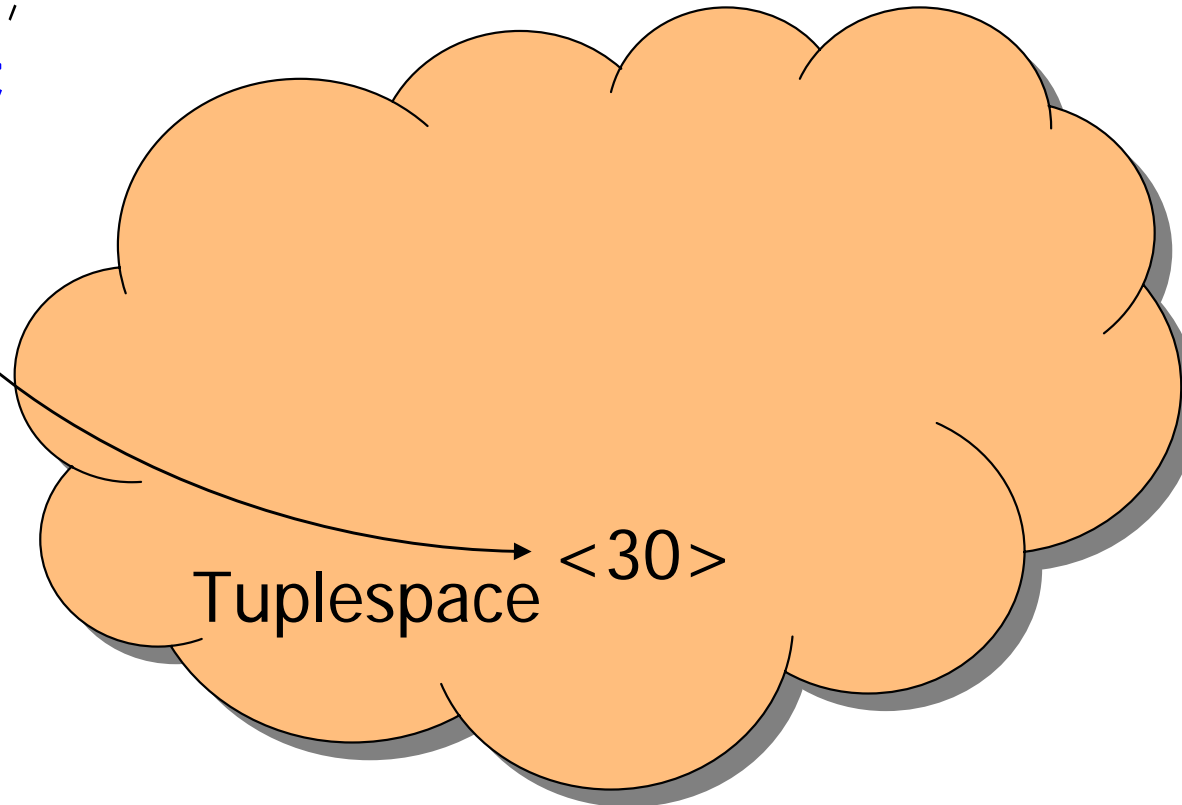
```
out(10,20);  
in(?result);
```



```
in(?a,?b);  
out(a+b);
```

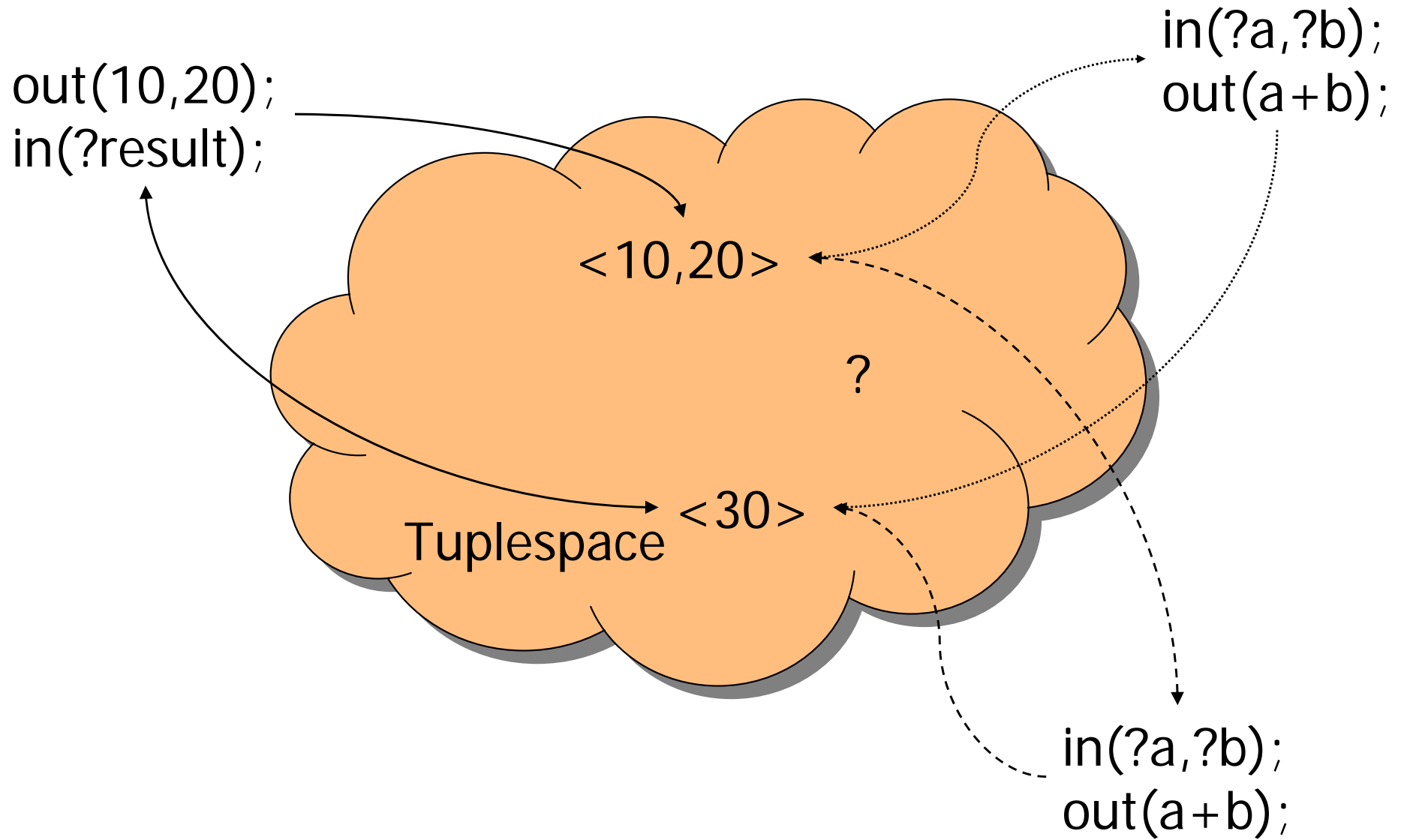
Linda: Indirect interaction

```
out(10,20);  
in(?result);
```



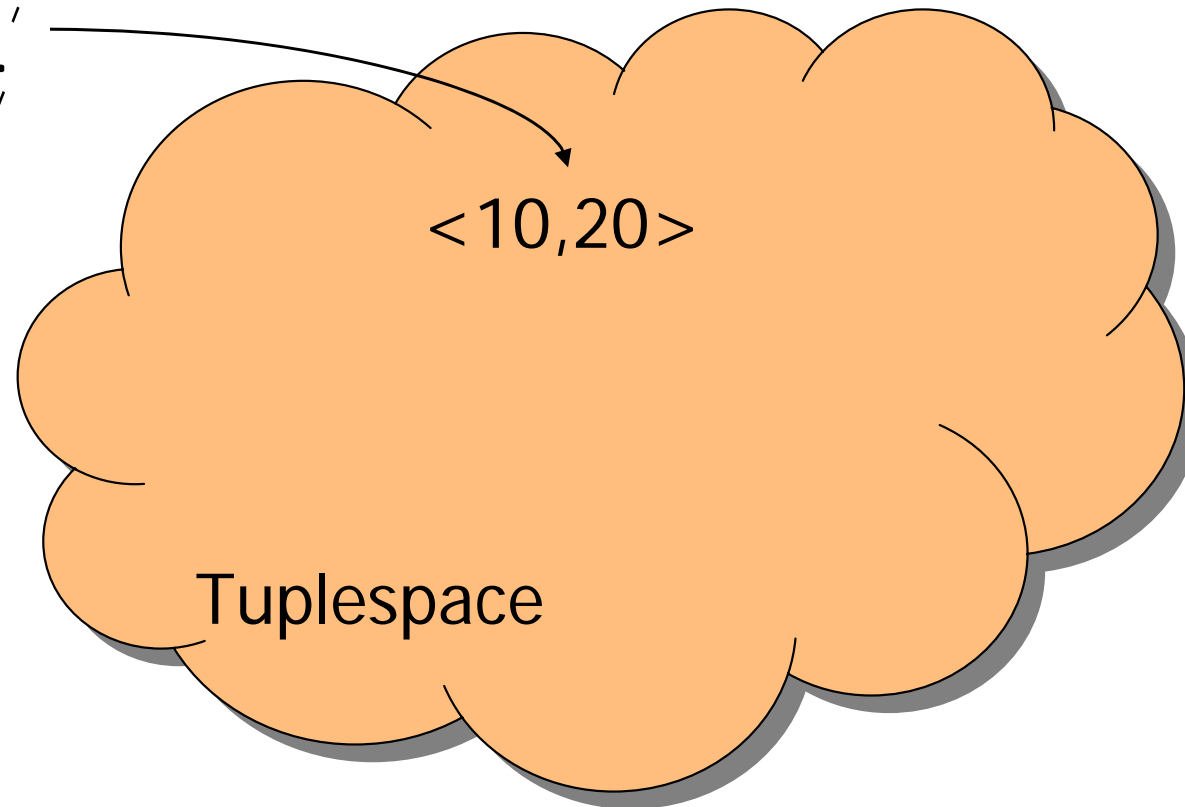
```
in(?a,?b);  
out(a+b);
```

Linda: Anonymous interaction



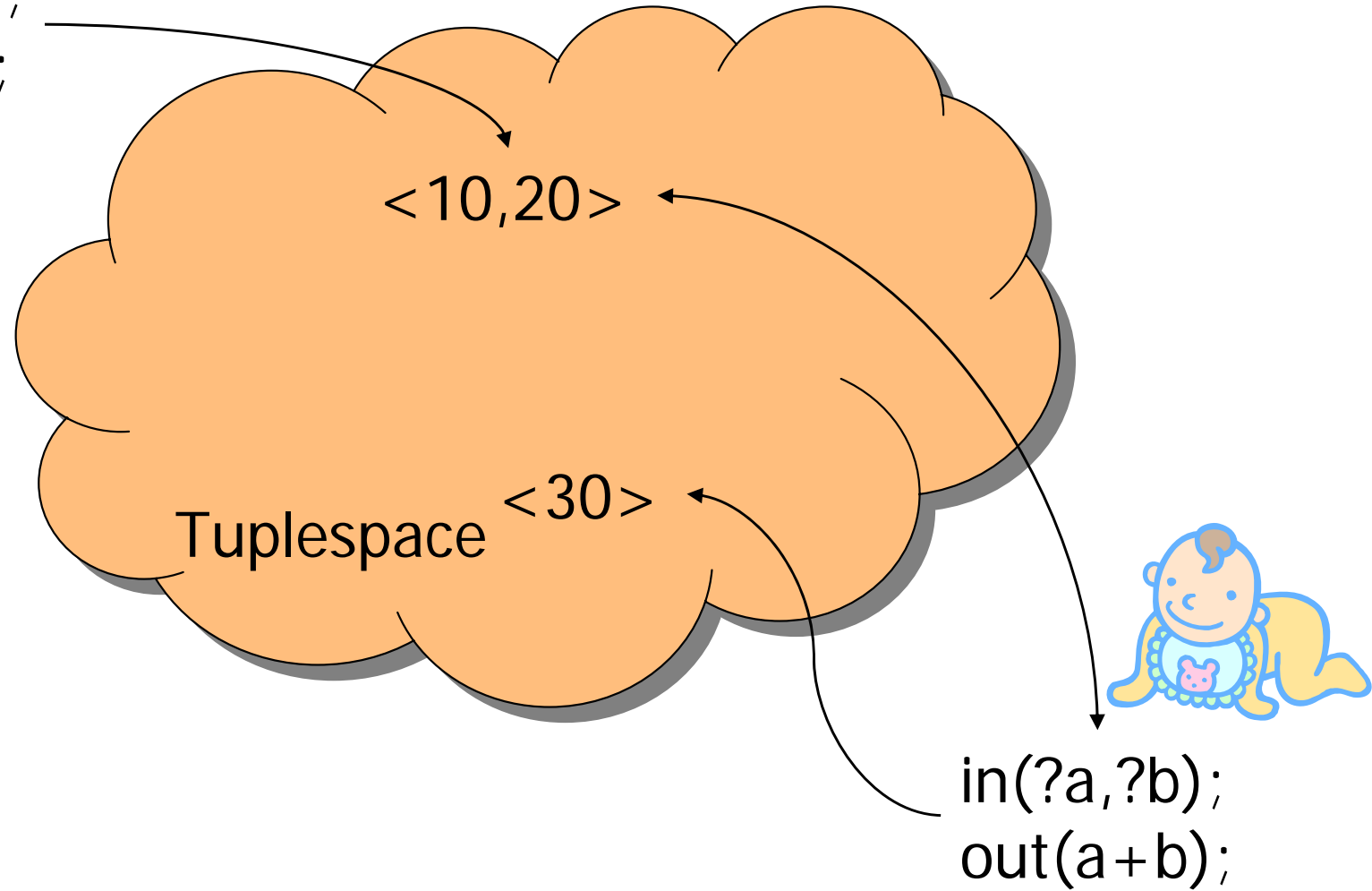
Linda: Non-overlapping lifetimes

```
out(10,20);  
in(?result);
```



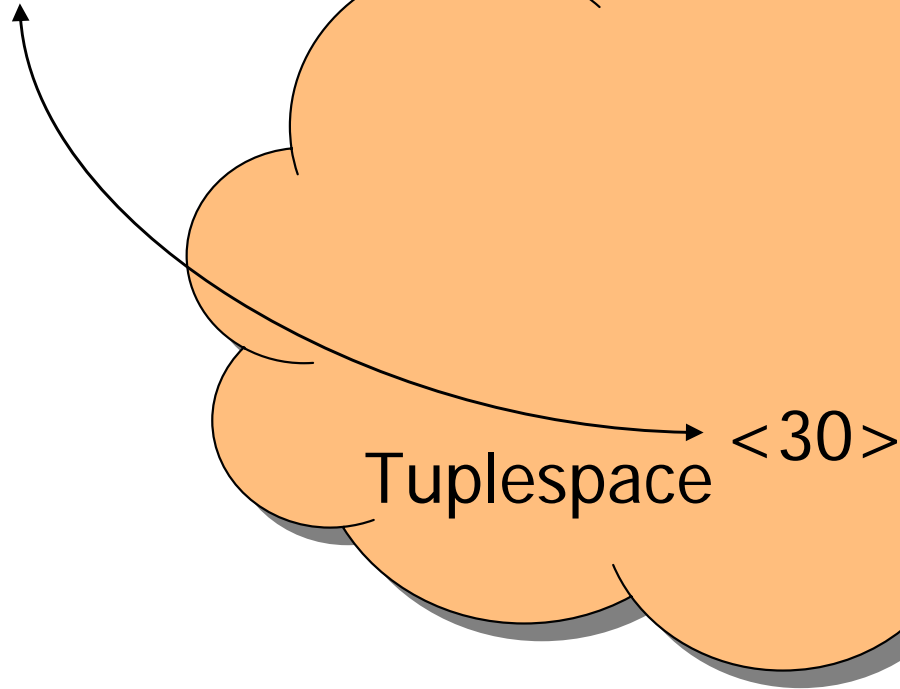
Linda: Non-overlapping lifetimes

out(10,20);
in(?result);

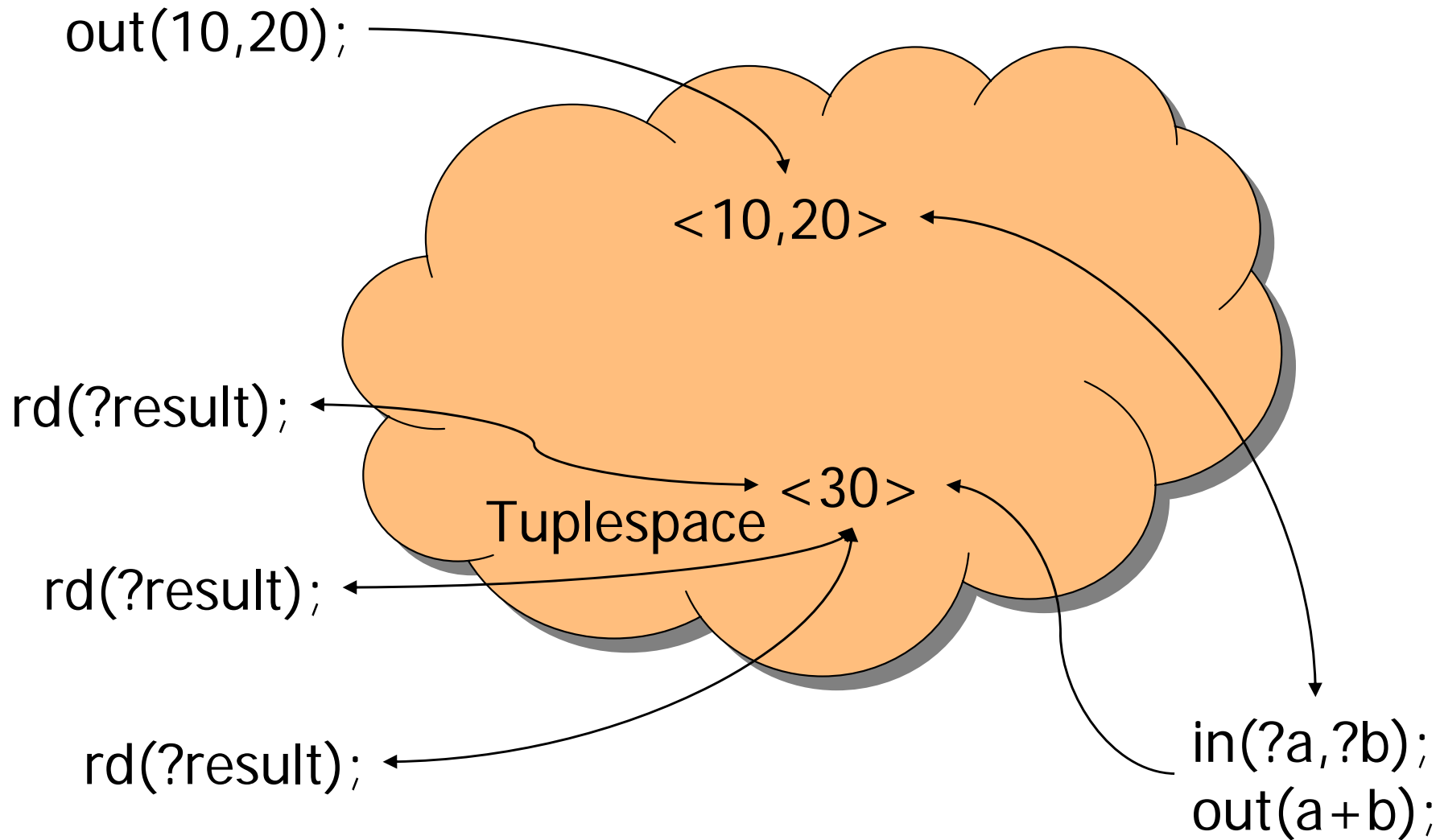


Linda: Non-overlapping lifetimes

```
out(10,20);  
in(?result);
```



Linda: Multiple party interaction



Idea: Semantic Web Spaces

- Extend Linda model so that
 - RDF triples (RDFSpaces)
 - OWL constructs (OWLSpaces)are
 - represented as first class objects in spaces
 - accessed by primitives
 - matched according to
 - RDF(S)/OWL (template is some other first class object)
 - RDQL/RQL/OWL-QL/SPARQL (template is a query)
- Extend space further so that
 - Rules type a space by limiting interactions (RuleSpaces)
 - Reasoners add proofs to the space (ReasonSpaces)
 - Agents exist in an active space and interact via trusted services (TrustSpaces)

- It is not that simple. We need to extend Linda:
 - Everything in the Semantic Web can be referenced by a URI. Linda-Tuples need a reference mechanism
 - Semantic Web information is typed. Linda-Tuples need richer typing than just the core data types
 - Data models such as XML and RDF/XML permit the nesting of elements within a single document. Linda-Tuples need nesting and means to express containedness
 - Semantic Web vocabularies containing the same terms can be kept separate using the namespaces mechanism. Linda Tuplespaces need a separation mechanism.

New Tuple types

- All RDF Statements are triples, thus regular tuples
<subject, predicate, object>
- Triples have to be referenced:
<subject, predicate, object, ID>
- ID has to be system determined to be unique
- Example (using QNames instead of full URIs):
A certain book has a certain ISBN number:
<my:novel [books:Book], books:has-isbn,
isbn:508-1762-443 [books:Isbn], my:bookisbn>
- RDF tuple with a blank node could look like this:
<my:person [foaf:Person], vcard:address,
_:80bg42c [ts:BlankNode], my:personsaddress>

New Tuple types

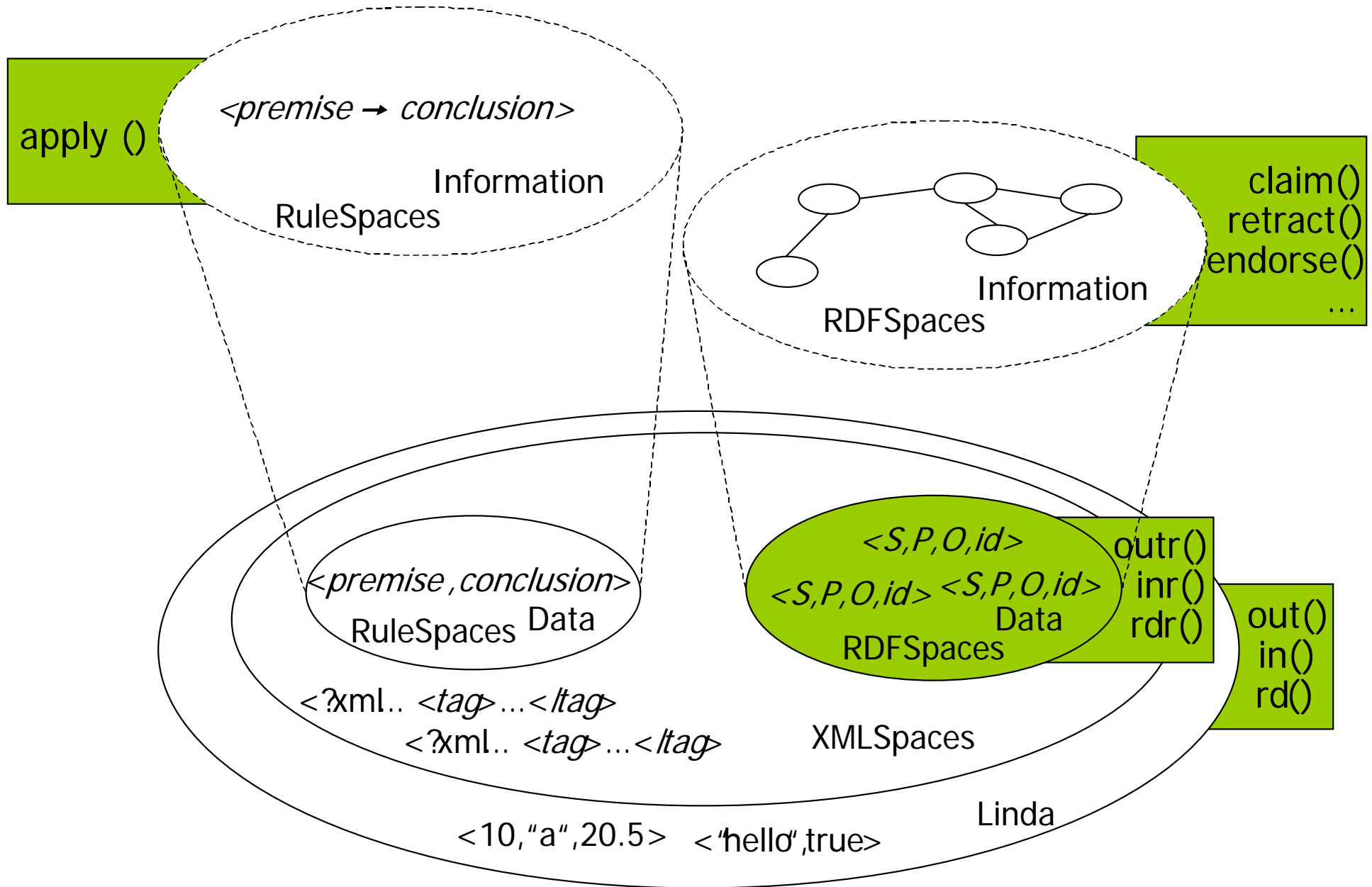
- Reification:
State my belief that one of my friends is currently in Australia:

```
<my:person [foaf:Person], trust:believes,  
  <my:friend3 [foaf:Person], loc:is-located-in,  
    country:Australia [country:Country],  
    my:statementId> [rdf:Statement],  
my:claim>
```

New coordination primitives

- Linda: Data is subject to operations
 - It is put into the space with **out** and *exists* there
 - It is removed from the space with **in**, it *no longer exists*
- Semantic Web: Information is subject to operations
 - RDF triple represents a *true* statement
 - When put into the space it exists and is true
 - When removed, it no longer exists but *is still true!*
 - removing is not destructive like the **in**-operation
 - removing does not mean that its negation is true
 - removing would rather mean that truth is unknown
- Two views in Semantic Web Spaces
 - Data view: uninterpreted triples
 - Information view: Interpreted information
- Different primitives work on different views

Data and information views



- Data view primitives
 - out: tuple -> boolean.
 - rd: template -> tuple.
 - in: template -> tuple.
- Typed data view primitives
 - outr: (s,p,o) -> boolean.
 - rdr: (s,p,o,id) -> tuple.
 - inr: (s,p,o,id) -> tuple.
- IDs are assigned internally by the system and are retrieved using special matching templates.

Information view: Making a statement

- claim: (s,p,o,id) -> boolean.
 - a single RDFS Triple
- claim: (Subspace) -> boolean.
 - a Subspace that contains one or more tuples.
 - claim multiple tuples in one operation
 - make claims which contain blank nodes. Within a Subspace, a blank node with the same identifier will be considered as being the same blank node when tuples are added into the space
- A claim carries a truth value, i.e. it is making a statement about something that it purports to be true
- After a claim the RDF tuple or set of RDF tuples
 - Exist in the data view
 - Its truth is a property of the tuple at the information view level

- endorse: (s,p,o,id) -> Subspace.
 - A tuple matching the given template is considered 'endorsed' by the information view, i.e. that it has been found to be consistent with current ontological information
 - The match is returned as a subspace
 - This subspace may contain a single matching tuple, however in the case of blank nodes the linking tuples are included in the response

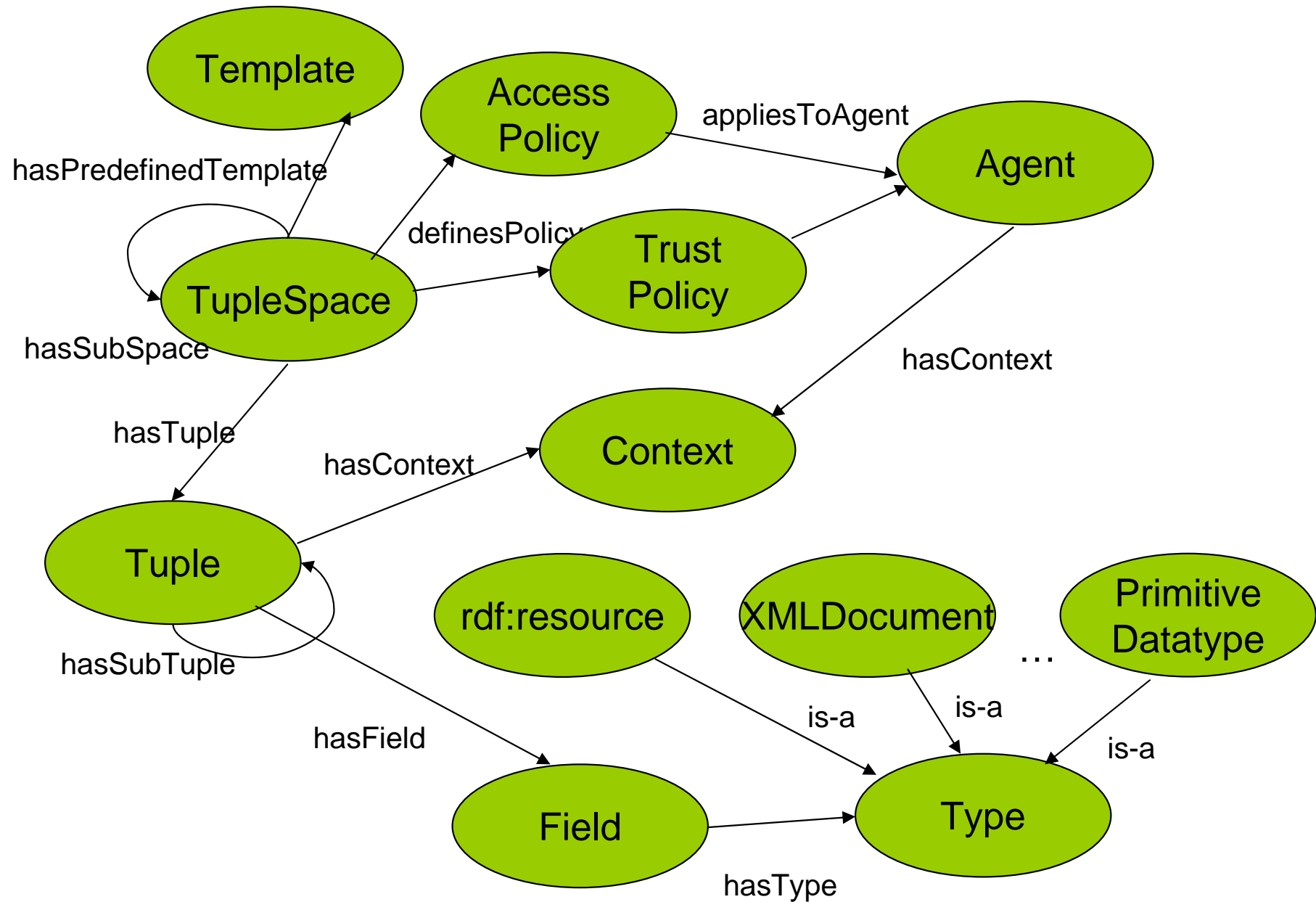
- excerpt: $(s,p,o) \rightarrow$ Context.
 - a version of the copy-collect primitive [RW98] that acts on the information view
 - It works within contexts: a context is created by the system into which all matching tuples are copied [TPBN05].
 - A reference to this context is passed to the client who is given alone the right to access the context. The client can then make destructive reads in that context i.e. $\text{retract}(*, *, *, *)$ to remove all of the tuples.
 - When the context is empty the system destroys the context
- Context:
 - A partial view on the tuplespace
 - A tuple can exist in multiple contexts but only one space

- retract: (s,p,o,id) -> Subspace.
 - removing a statement from the set without denying its truthfulness
 - if a matching tuple is found, replaces its subject, predicate and object in the information view with bottom-element _ rather than removing it completely
 - `<my:novel [books:Book], books:has-isbn, isbn:508-1762-443 [books:Isbn], my:bookisbn>`
`<_, _, _, my:bookisbn>`
 - The tuple remains in the data view
The reference continues to exist but the assertion that the reference makes is lost or forgotten
- As a result, all inferrable tuples from that retracted tuple must also be lost

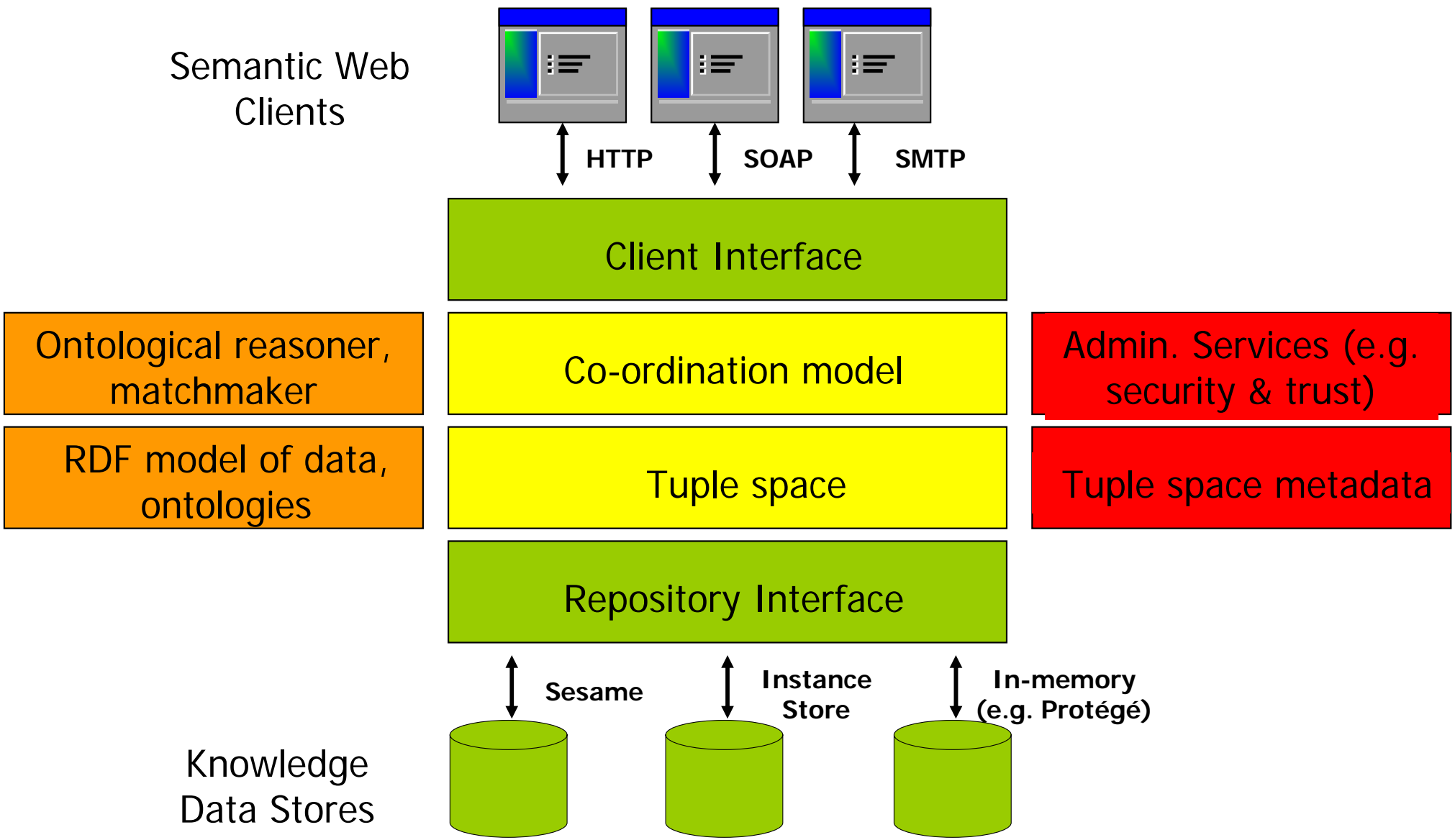
Templates

- Data view: similar to Linda:
 - `in(?,books:has-isbn,?)`
- Information view: Interpret data
 - In RDFSpaces: According to RDF semantics
 - `endorse(book,title,?x)`
means
`SELECT ?x WHERE(book,title,?x)`
 - Use subsumption
 - C1 subClassOf C2: $C1 \sqsubseteq C2$
 - P1 subPropertyOf P2: $P1 \sqsubseteq P2$
 - In matching, any T1 in a template is considered a match with any T2 in a tuple iff $T1 \sqsubseteq T2$
 - `endorse(media,title,?x)` will return a book or a CD if `book subClassOf media` and `cd subClassOf media`

Tuple Space Ontology



Implementation idea



Semantic Web Spaces

- Semantic Web Spaces is a good alternative to traditional storage of Semantic Web information
 - decentralized
 - distributed
 - associative high-level access
- Semantic Web Spaces is a good alternative to traditional interaction on the Semantic Web
 - Uncoupled in space and time
 - simple
 - superior to client/server interaction

Outlook

- Further development of the idea in the upcoming EU funded TripCom project
- The technical report
Robert Tolksdorf, Elena Paslaru Bontas, Lyndon J B Nixon. *A Conceptual Model for Semantic Web Spaces*. Technical Report B-05-14, FU Berlin, Institut für Informatik, 2005. At <http://www.inf.fu-berlin.de/inst/pubs/tr-b-05-15.abstract.html> gives a use case
- The technical report
Elena Paslaru Bontas, Lyndon J B Nixon, Robert Tolksdorf. *Using Semantic Web Spaces to realize Ontology Repositories*. Technical Report B-05-15, FU Berlin, Institut für Informatik, 2005. At <http://www.inf.fu-berlin.de/inst/pubs/tr-b-05-14.abstract.html> gives all further details eg. on blank nodes and contexts

Tripcom

IST-4-027324-STP, Start: April 2006

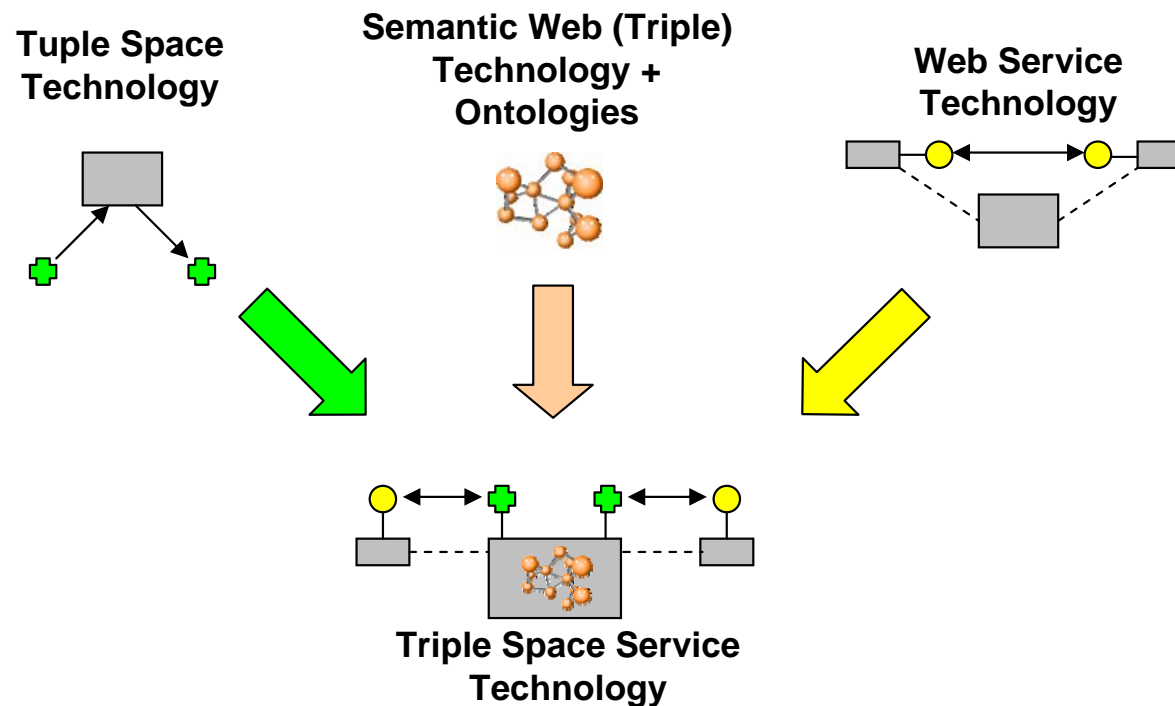
Project Presentation



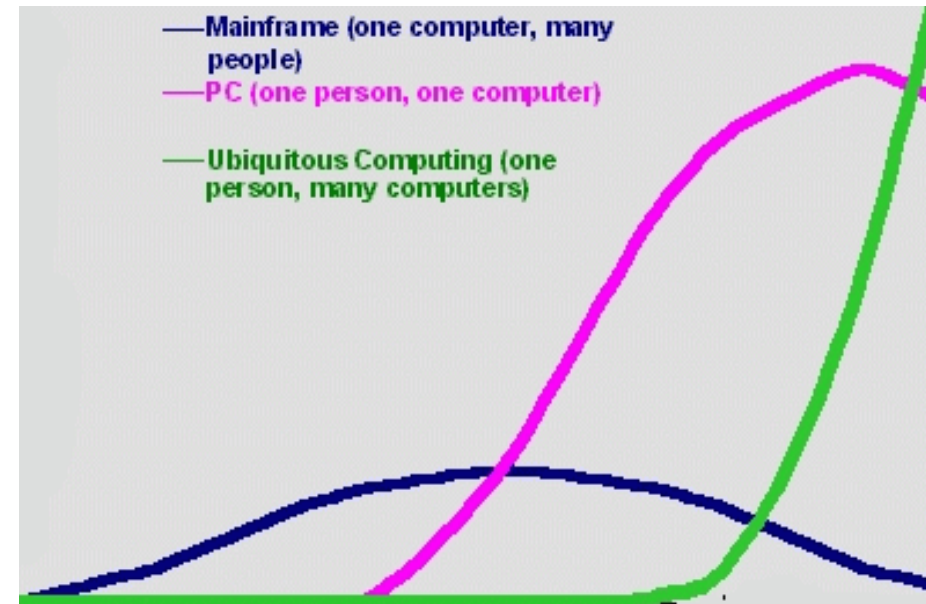
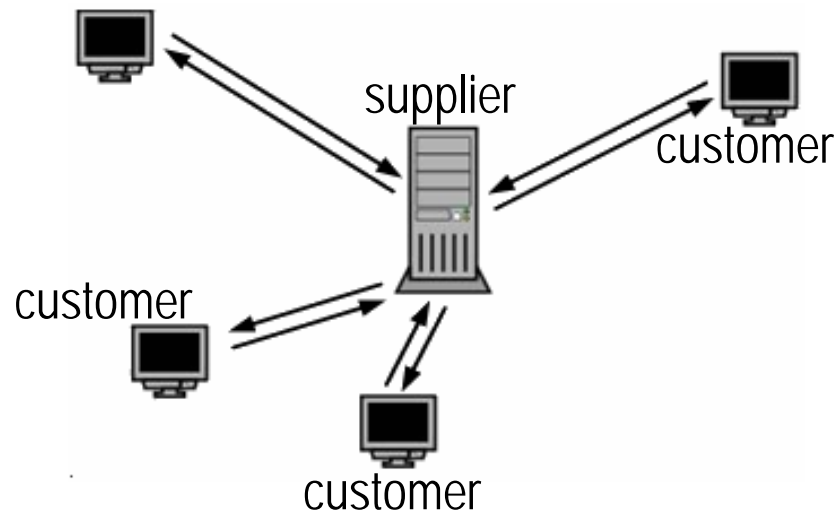
Berlin Meeting, 17 January 2007



Triple Spaces provide a communication paradigm for *anonymous, asynchronous* information exchange that ensures the *persistency* and *unique identification* of the communicated *semantic, semi-structured* data.



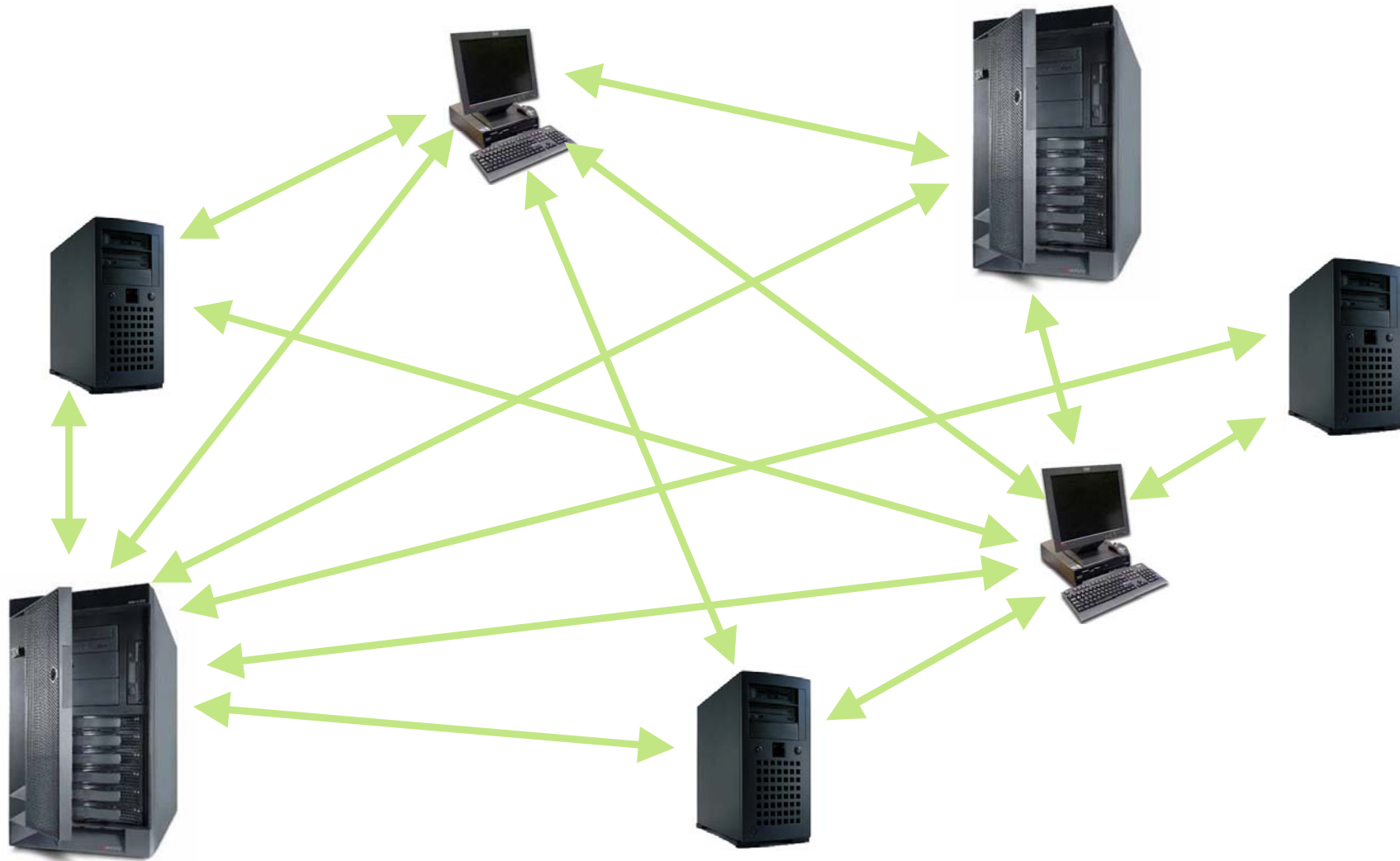
Computers interaction (since all times) is dominated by message exchange based communication.



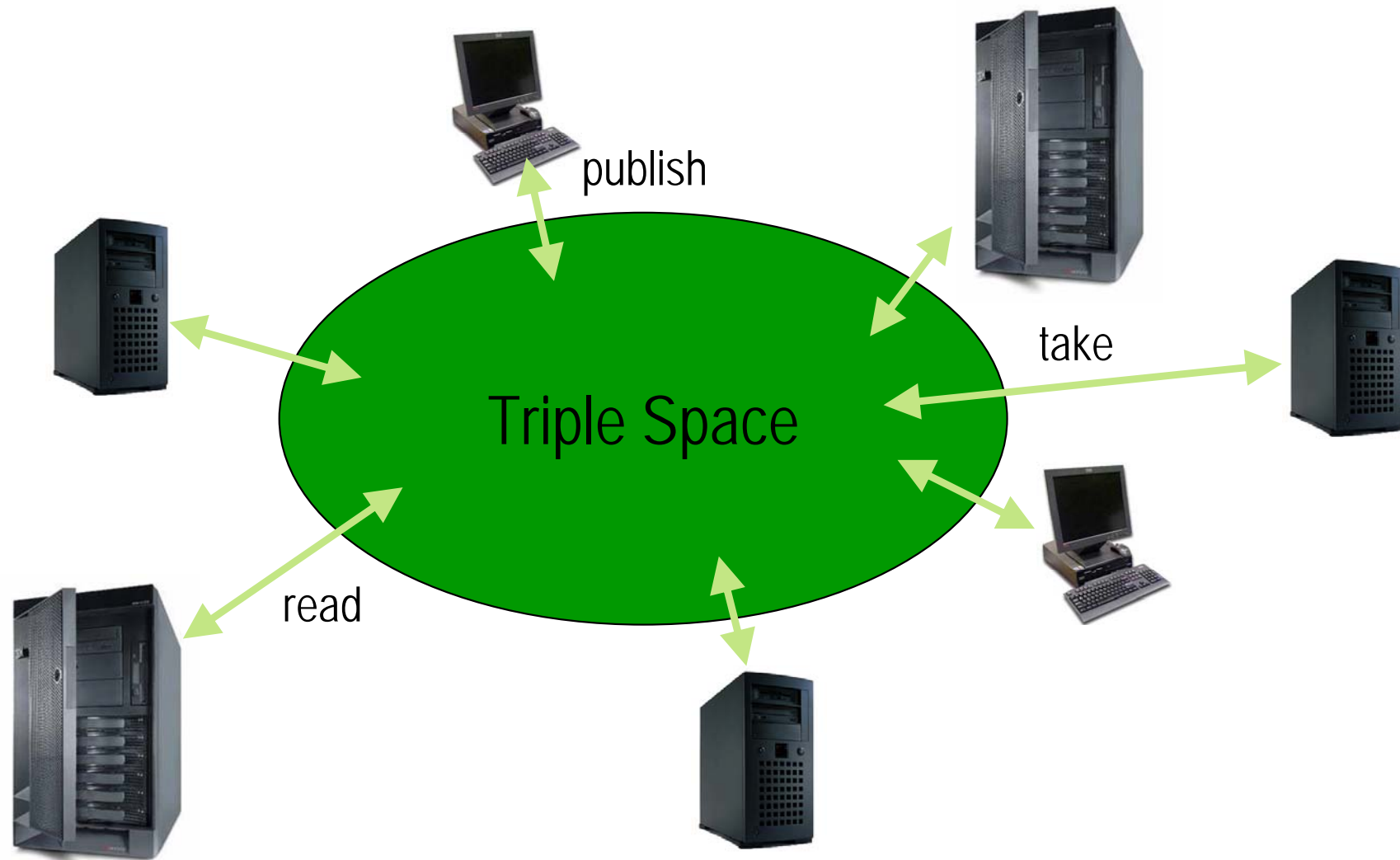
Why?

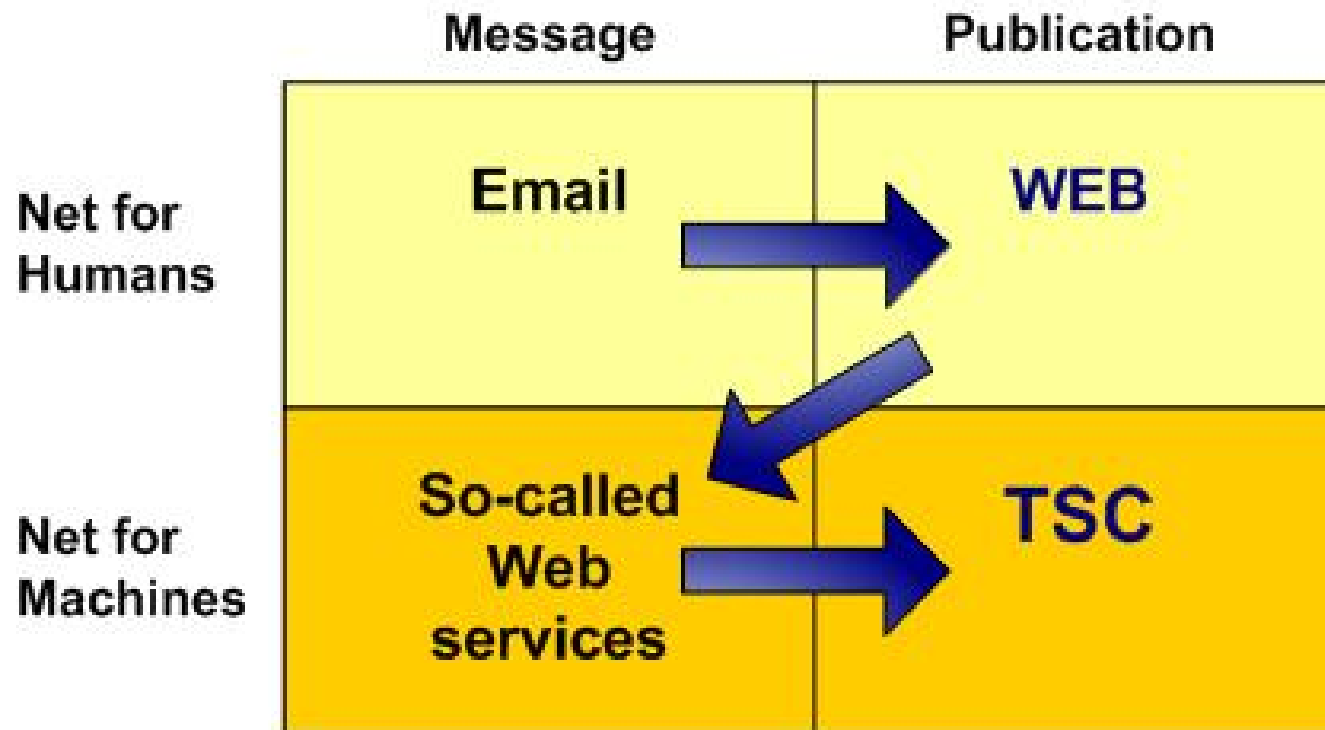
It implies unnecessary complexity on (machine to machine) interaction. The WWW only scaled to its current size because it uses persistent publication as the communication paradigm.

Communication (1)



Communication (2)

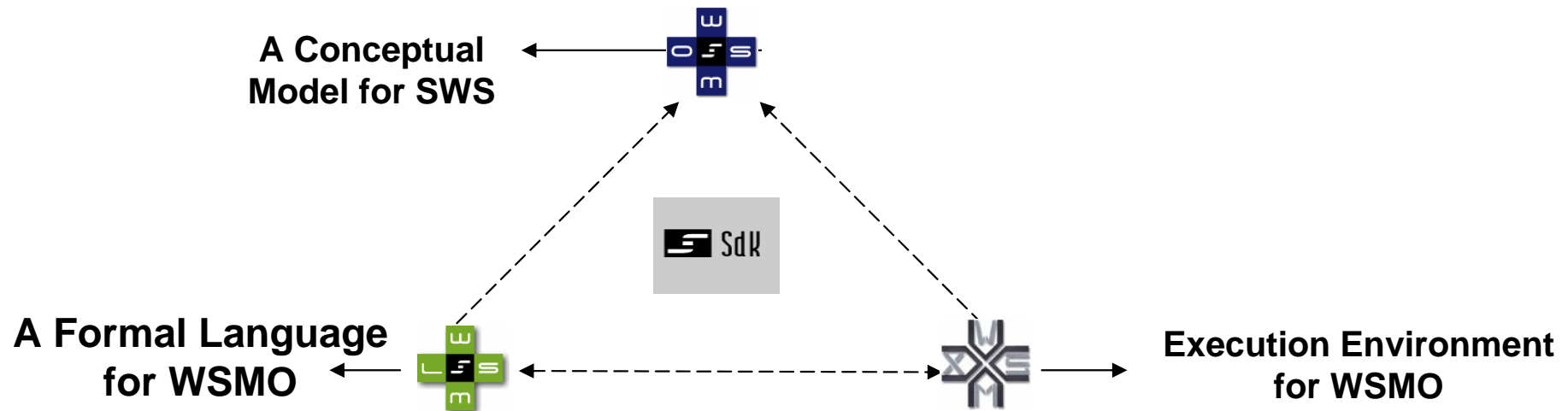




"Triple Space may become the web for machines as the web based on HTML became the Web for humans"

Semantic Web + Web Services

Semantic Web Services



Semantic Web + Tuple Space

Triple Space



Triple Space Computing (TSC) funded by FFG Austria

D. Fensel: *Triple-space computing: Semantic Web Services based on persistent publication of information*, Proc. of IFIP Int'l Conf. on Intelligence in Communication Systems, Bangkok, Thailand, November 2004: 43-53.

R. Krummenacher, M. Hepp, A. Polleres, Ch. Bussler, and D. Fensel: *WWW or What is Wrong with Web Services*, Proc. 3rd European Conf. on Web Services ECOWS2005, Växjö, Sweden, November 2005: 235-243.



Semantic Web Services + Triple Space

TripCom

- Strengthen expertise in the fields of:
 - Semantic Web (RDF)
 - Web services (WS-* Standards)
 - Linda/Tuple Spaces (extensions towards Semantic Web)
 - Applications in Industry and Public Sector



Facts & Figures



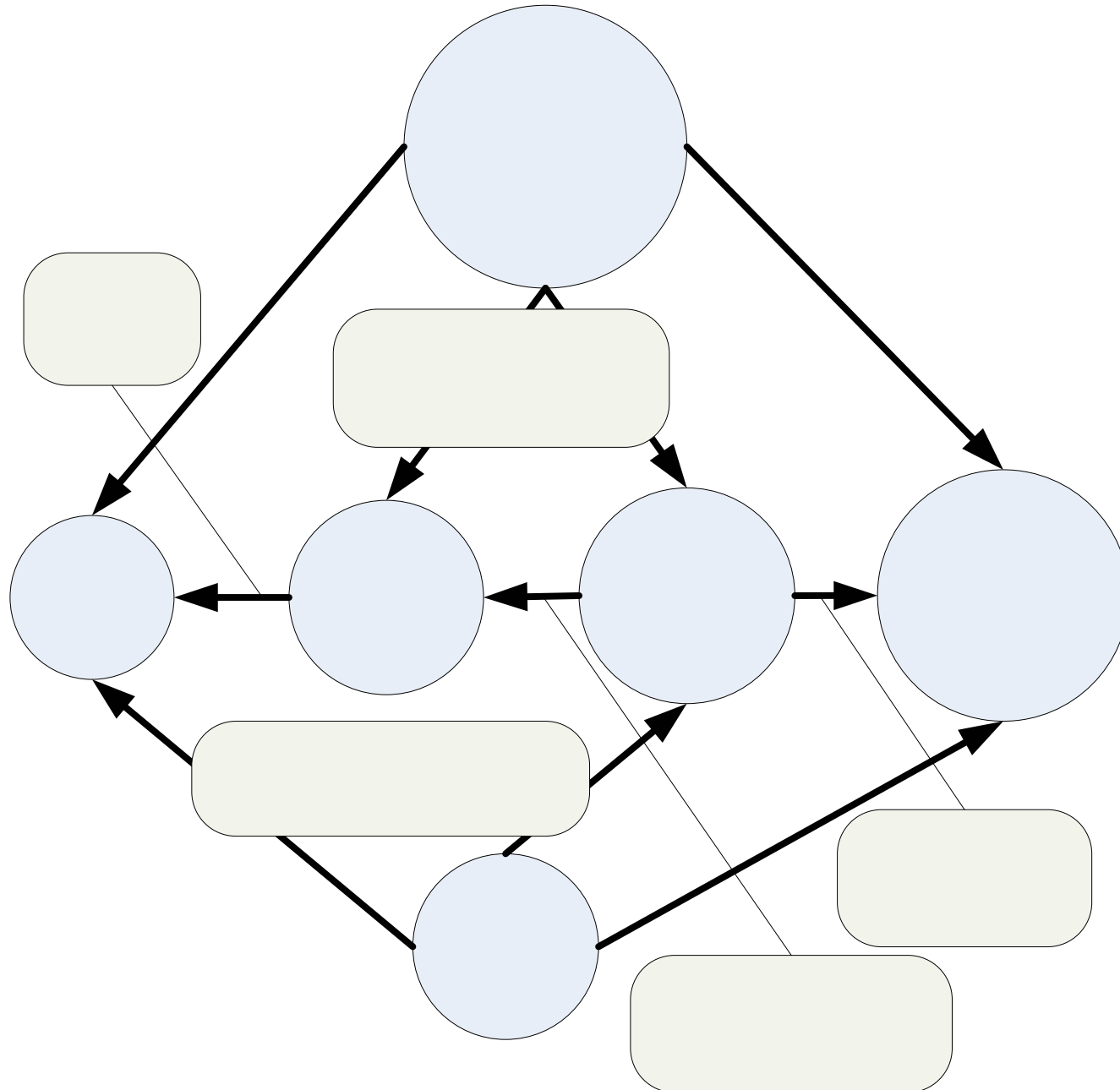
Start date: April 2006
End date: March 2009
Duration: 36 month

Number of partners: 9
Number of countries: 7 (AUT, BUL, ESP, FIN, GER, IRL, ITA)

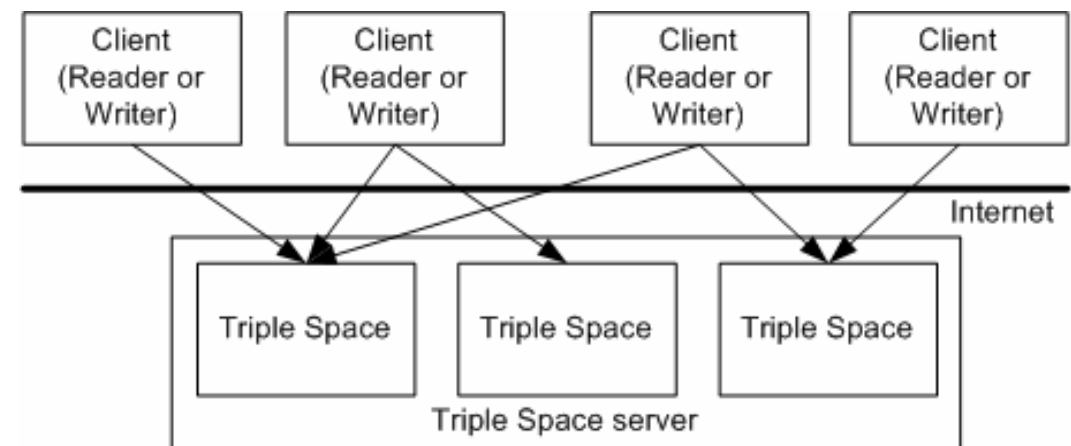
Number of work packages: 11
Number of deliverables: 46
Number of person*month: 623

Budget total: €4'661'110.-
EC funding: €3'650'000.-





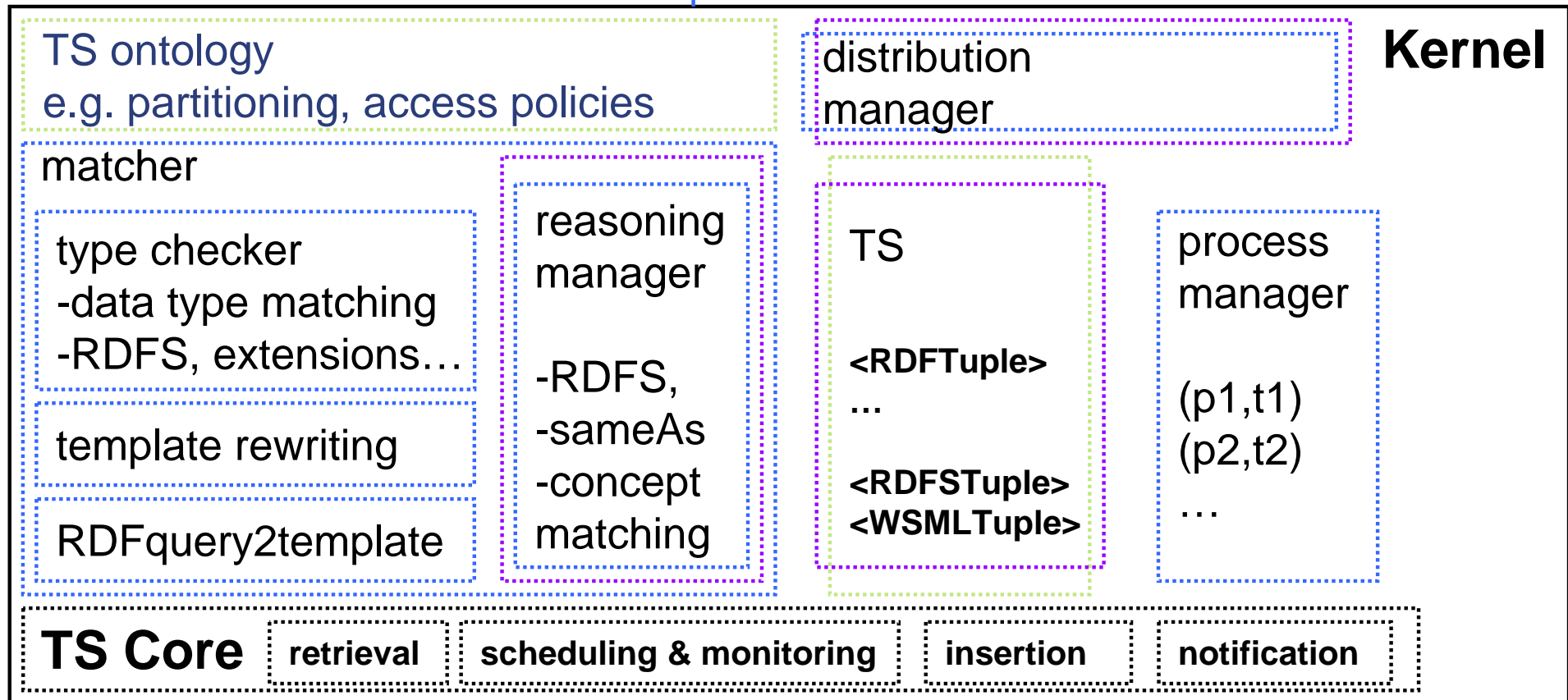
- Triples
 - The semantic data exchanged or shared
 - Based on RDF data model of <subject, predicate, object>
- Triple Space
 - The space used to exchange or share triples
- Triple Space server
 - The infrastructure / server hosting Triple Spaces



Triple Space Architecture (WP6)

Kernel

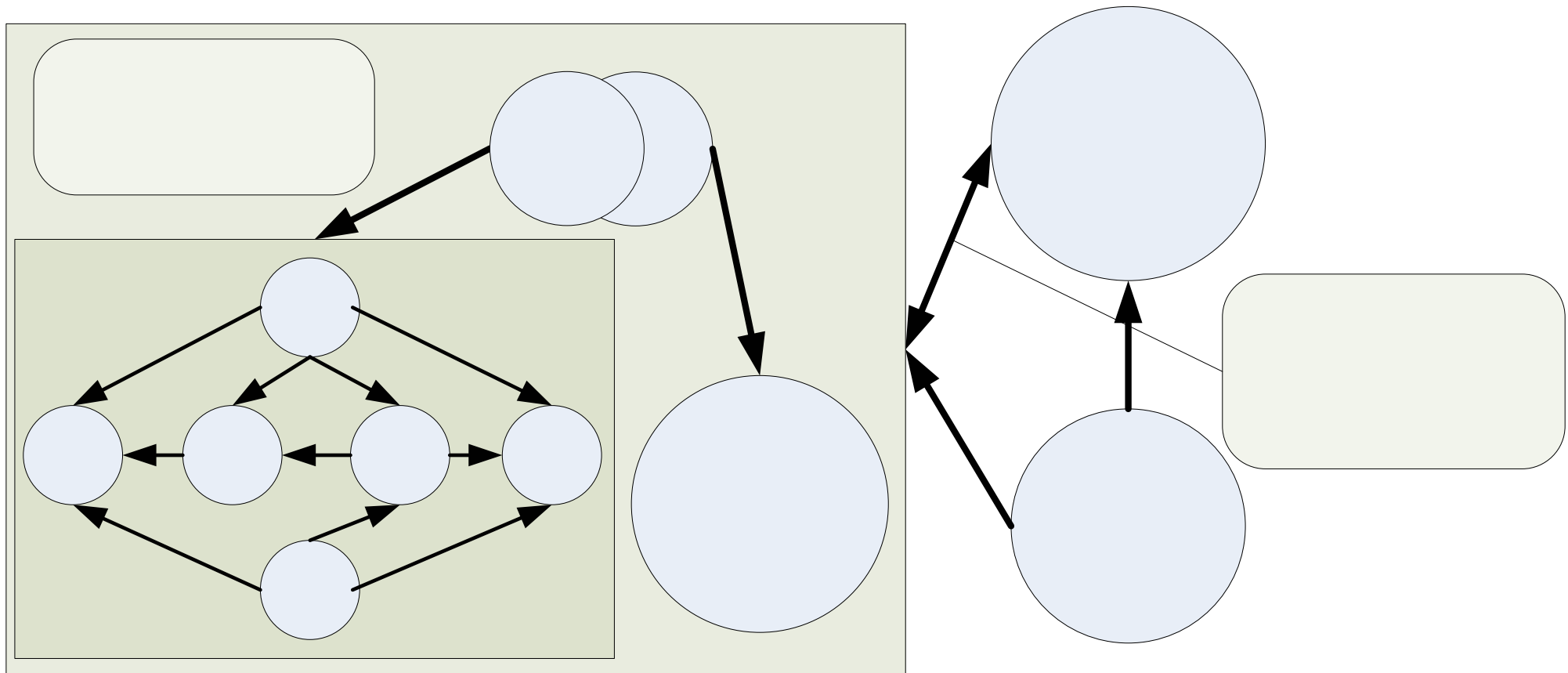
Triple Space Kernel API



Kernel

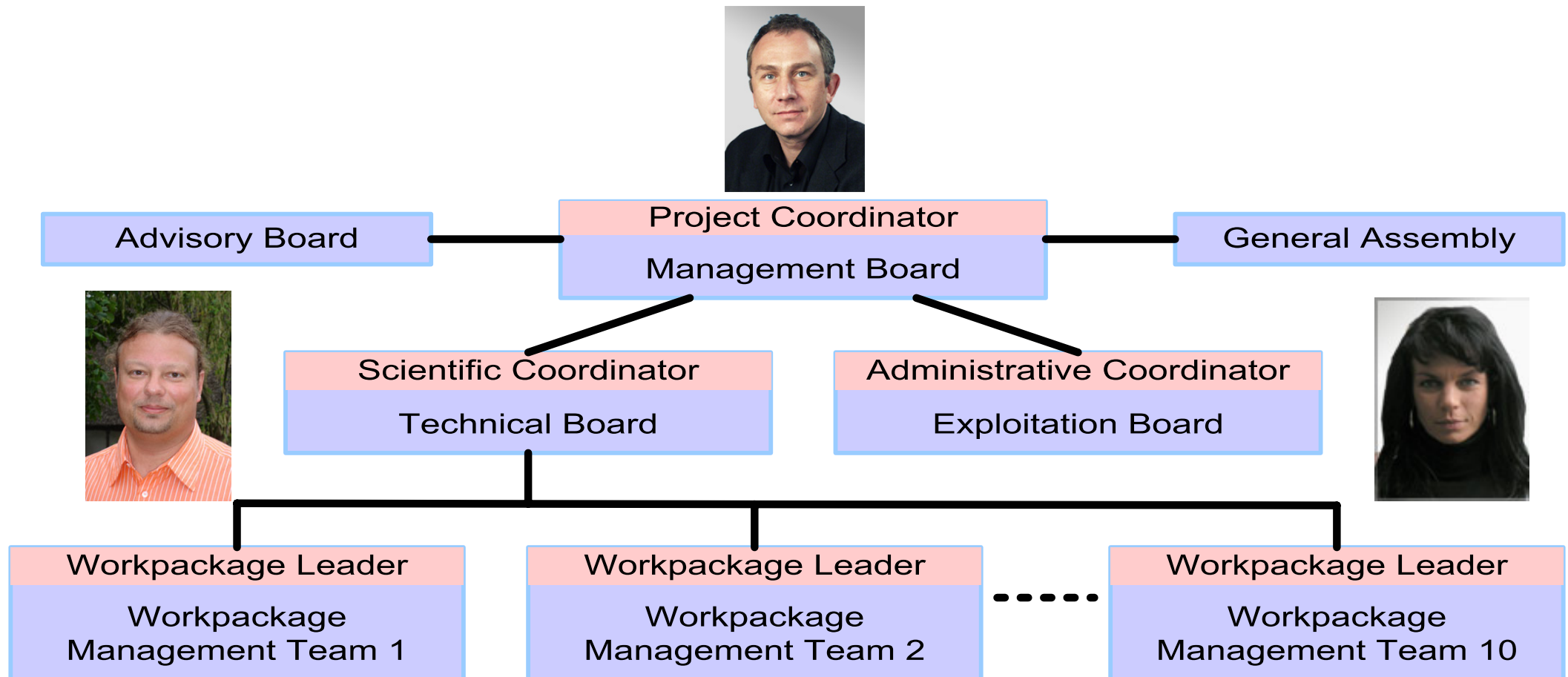
Triple Space API

Web Service



- Aim:
Realise Triple Space as a global infrastructure for machine communication of semantic data based on the Web principles of persistent publication and simple interaction („RESTfulness“)
- Specific use cases in the project:
 - Extend Semantic Web Service communication (WP4)
 - Enhance EAI with semantics and coordination (WP7, 8a)
 - Enable the infrastructure necessary to support a European Patient Summary (WP8b)

Structure and Key Persons



Effective management of quality is of fundamental importance to the success of the project.

- TripCom has implemented a Quality Assurance Procedure which ensures that every deliverable has to undergo a three-step quality check before it is submitted to the EC
- The Quality Assurance Procedure starts exactly 1 month before the final EC submission date of the respective deliverable
- Reviewers involved in the Quality Assurance Procedure have been assigned for all deliverables due until M18. The complete list (Quality Assurance Responsibles) is available at <http://www.tripcom.org>

Quality assurance schedule

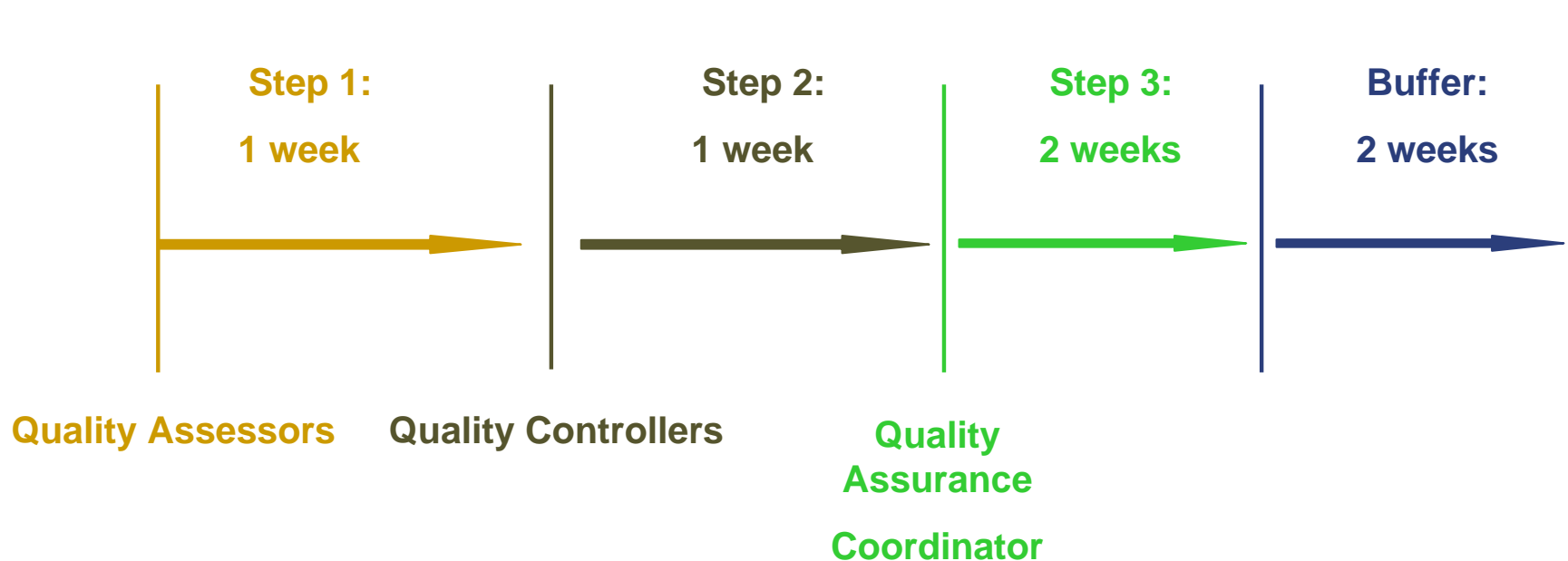
Any delay must be reported ASAP to the Administrative Coordinator who will handle the situation accordingly.

Deliverable

Deadline

(M3, M6, etc.)

**Submission
to
Commission
(up to 1 month
after deliverable
deadline)**



■ Quality Assessor

- A QA is a WP-external quality reviewer from one of the Consortium partners.
- A QA is assigned to every deliverable.

■ Quality Controller

- A QC is the leader of the WP or assigned from leader of the WP, in which the deliverable is produced.
- A QC is responsible for last and overall check of deliverables before they are submitted to Quality Assurance Coordinator.

■ Quality Assurance Coordinator

- The QAC is the Technical Coordinator of TripCom.
- The QAC is responsible for last and overall check of all deliverables before they are submitted to EC.



<http://www.sdk-cluster.org>



SEKT (Semantically-Enabled Knowledge Technologies)
<http://sekt.semanticweb.org/>



Data, Information and Process Integration
with Semantic Web Services

DIP (Data, Information and
Process with Semantic Web Services)

<http://www.nextwebgeneration.org/projects/dip/>



Knowledge Web

<http://knowledgeweb.semanticweb.org/>

Other Related efforts/projects:



Super <http://www.ip-super.org>

Cocoon <http://www.cocoon-health.com/>

Infrawebs <http://www.infrawebs.org/>

Target Standardization Bodies:



World Wide Web consortium (W3C)



Organization for the Advancement of Structured
Information Standards (OASIS)

1. Enhance existing products of the commercial partners of TripCom, e.g. Telefonica
 1. Leverage TripCom implementations of semantically enhanced service interaction patterns
 2. Target BPEL Standardization
 3. Incubator programme for "Serviceoutsource.com"
 4. Target public sector as early adopter
2. Identify new business possibilities / raise venture capital for new start-ups, e.g. CEFRIEL
3. Address other companies via our dissemination strategy



Topic Maps

Teile nach Steve Pepper: The TAO of Topic Maps
(<http://www.ontopia.net/topicmaps/materials/tao.html#d0e632>)

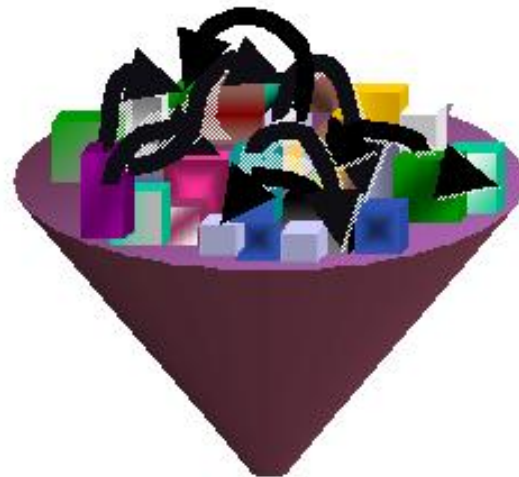
Topic Navigation Maps...

- ... Repräsentation von Semantik von Informationen
- Information technology - SGML Applications - Topic Maps
 - Ziel: Adressierbare Information semantisch zugänglich machen
 - Status: ISO/IEC 13250:2000
 - Quelle: <http://www.topicmaps.net/>,
<http://www.y12.doe.gov/sgml/sc34/document/0129.pdf>
- XML Topic Maps (XTM) 1.0
 - Ziel: XML-basiertes Austauschformat für Topic Maps
 - Status: TopicMaps.Org
 - Quelle: <http://www.topicmaps.org/xtm/1.0/>

Why Topic Maps?

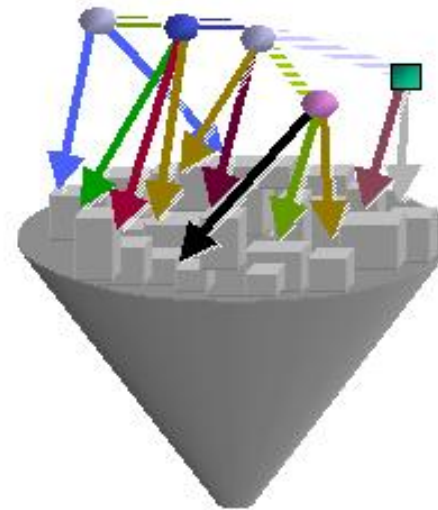
Before

Infoglut



After

Information / Knowledge Management



InfoLoom, Inc.
and
Coolheads
Consulting

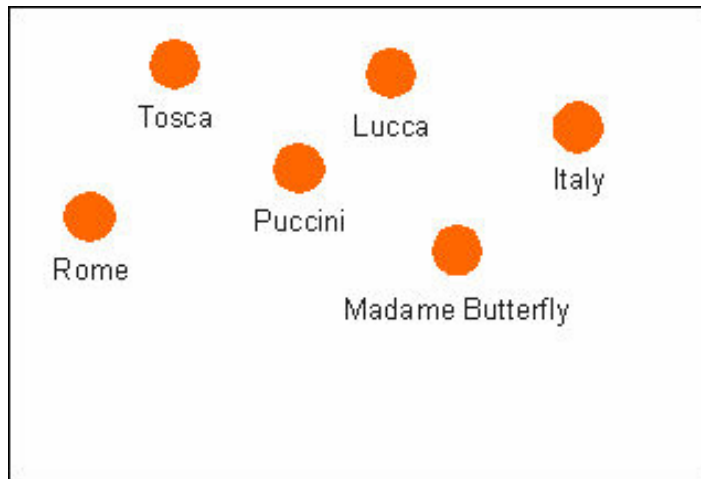
© 2001, InfoLoom, Inc. and Coolheads Consulting. All rights reserved.

- Bücher können durch einen Index erschlossen werden
 - *Cavalleria Rusticana*, 71, **203-204**
The Girl of the Golden West, see *La fanciulla del West*
Leoncavallo, Ruggiero *I Pagliacci*, 71-72, 122, **247-249**,
326
Madama Butterfly, 70-71, **234-236**, 326
- Spiegeln Struktur des Buches / den enthaltenen Wissens wider
- Volltextindex:
 - Indexiert alle Worte
 - Gleiches Thema mit mehreren Bezeichnungen unerkannt (Synonyme)
 - Mehrfache thematische Bedeutung eines Begriffs unerkannt (Homonyme)
- Topic Maps:
 - Indexing + Bibliothekswesen + Wissensrepräsentation

- Buchindex / Register enthält
 - geordnete Liste von Themen (**T**opics)
 - Beziehungen zwischen Themen (**A**ssociations)
 - Verweise auf deren Auftreten im Buch (**O**ccurrences)
- Buchindex / Register benutzt
 - Typographische Konventionen zur Typisierung von occurrences
 - Weitere Typisierungen („definiert bei“, „benutzt bei“)
 - siehe / siehe auch Verweise auf andere Themen
 - Geschachtelte Einträge
 - Markierungen bei Homonymen
 - Tosca (opera)
 - Tosca (character)

Topic

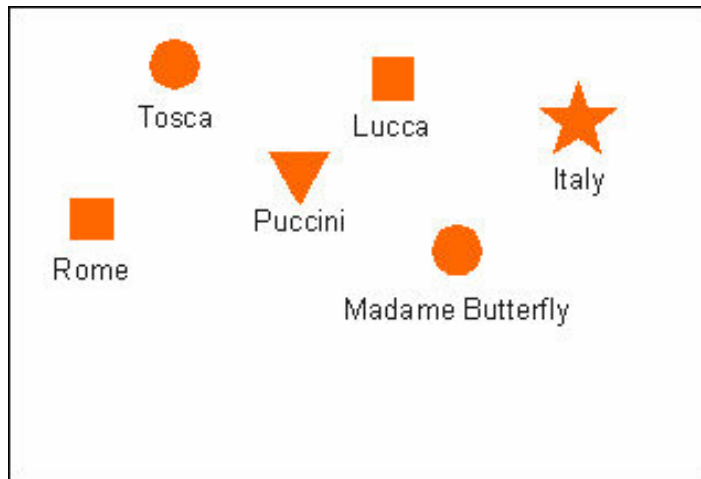
- Topic ist „etwas“
 - Person, Konzept, Ding



- Topic repräsentiert ein Subjekt in der realen Welt
- 1:1 Beziehung zwischen Topic und Subjekt

Topic Typen

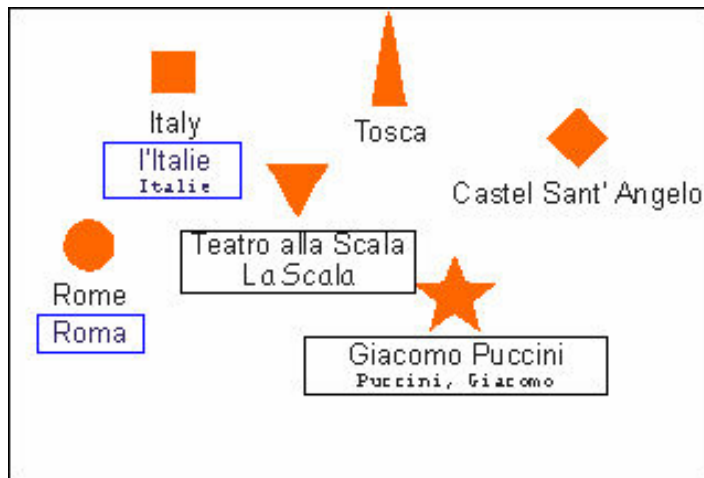
- Topics können nach ihrer Art gruppiert werden
- Jedes Topic kann 0..n Typen angehören



- Topic Typen sind selber Topics

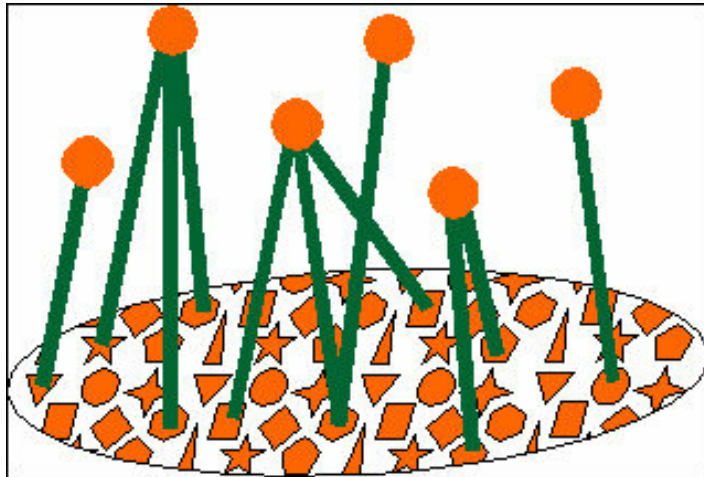
Topic Namen

- Topics haben normalerweise Namen
- Einige haben keine Namen („siehe Seite 38“)
- Einige haben viele Namen



Occurrences

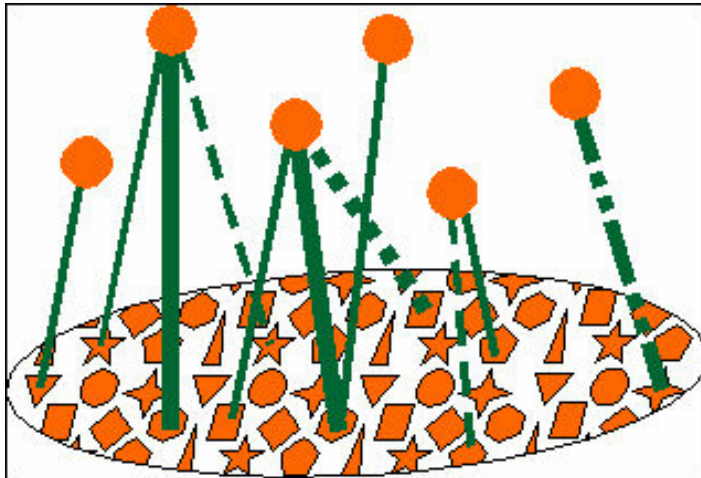
- Einem Thema sind Occurrences zugeordnet
- Occurrences: Für das Thema relevante Ressourcen
 - Bücher, Artikel, Photo, Nennung...



- Occurrences sind extern zur Topic Map
- Werden referenziert, z.B. mit URL

Occurrence roles

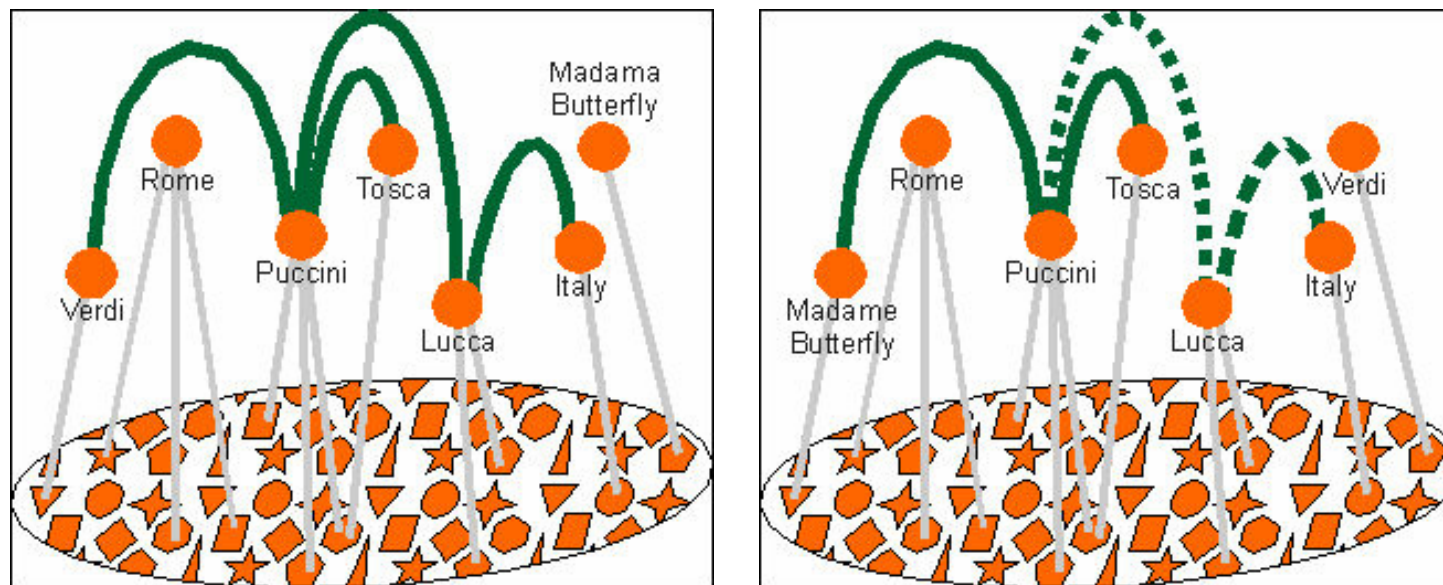
- Typisieren die Occurrence



- Occurrence role: Kürzel
- Occurrence type: Topic

Associations

- Sind Beziehungen zwischen Topics
 - "Tosca was *written by* Puccini"
 - "Tosca *takes place in* Rome"
 - "Puccini was *born in* Lucca"
- Können typisiert sein



- Association role: Welche Rolle spielt Topic in Association

Subject Identifier

- In zwei unterschiedlichen Topic Maps können zwei Topics dasselbe Subjekt repräsentieren
- Bei Vereinigung muss 1:1 Beziehung erhalten bleiben
- Topic Map Lösung
 - Subject Indicator: Ressource, die eine eindeutige Klärung der Identität des Subjekts erlaubt
 - z.B. RFC Dokument mit Länderkürzeln
 - Da es sich um eine Ressource handelt, gibt es eine eindeutige Adresse (URL)
 - Zwei Topics, die einen oder mehrere SIs gemeinsam haben, sind semantisch äquivalent zu dem Topic das sich aus der Vereinigung ihrer Associations, names und Occurrences ergibt
- Published Subject Indicator: Gepflegte und veröffentlichte Subject Indicators (vgl. Ontologie)

Topic Navigation Maps Konzepte

- *Subject*: Etwas, das benannt oder wahrgenommen werden kann
- *Topic*: Repräsentation eines *Subject* in einer Topic Map
- *Name*: Bezeichnung eines Topics
- *Occurance*: Information, die in Bezug auf ein Subject wichtig ist
- *Association*: Beziehung zwischen Topics
- *Scope*: Kontext, in dem Namen, Occurances und Associations gültig sind
- *Topic Map*: Sammlung von Topics, Associations und Scopes
- *Merging*: Kombinieren zweier Topics oder Topic Maps

- Umschließendes Element: `<topi cMap>`
- Verweis auf ein anderes Topic: `<topi cRef>`:
`<topi cRef xlink:href=
"http://www.topicmaps.org/xtm/1.0/language.xtm#en" />`
- In `<topi c>`: Klassenbeziehung mit `<i nstanceOf>`:
`<topi c id="hamlet">
 <i nstanceOf>
 <topi cRef xlink:href="#play" />
 </i nstanceOf>
</topi c>`

- Verweis auf Subject: `<subjectIdentify>`
`<topic id="dk">`
 `<subjectIdentify>`
 `<subjectIndicatorRef xlink:href=`
 `"http://www.topcmaps.org/xtm/1.0/country.xtm#dk"/>`
 `</subjectIdentify>`
`</topic>`
- Verweise
 - Auf Resource (Subjekt ist durch sich selber gegeben):
 `<resourceRef>`
 - Auf anderes Topic: `<topicRef>`
 - Auf Subjektbeschreibung: `<subjectIndicatorRef>`

- Gültigkeitsbereiche durch `<scope>`:
`<scope>`
 `<subjectIndicatorRef xlink:href=`
 `"http://www.topicalmaps.org/xtm/1.0/language.xtm#en" />`
`</scope>`
- Namen von Topics:
`<topic id="denmark" >`
 `<baseName>`
 `<scope><topicRef xlink:href="#en" /></scope>`
 `<baseNameString>Denmark</baseNameString>`
 `</baseName>`
 `<baseName>`
 `<scope><topicRef xlink:href="#da" /></scope>`
 `<baseNameString>Danmark</baseNameString>`
 `</baseName>`
`</topic >`

- Varianten mit <variant>:

```
<topic id="shakespeare">
  <baseName>
    <baseNameString>William Shakespeare</baseNameString>
    <variant>
      <parameters>
        <topicRef xlink:href="#sort"/>
      </parameters>
      <variantName>
        <resourceData>shakespeare, William</resourceData>
      </variantName>
    </variant>
  </baseName>
</topic>
```

- Assoziationen:

```
<association id="will-wrote-hamlet">
  <instanceOf>
    <topicRef xlink:href="#written-by"/>
  </instanceOf>
  <member>
    <roleSpec>
      <topicRef xlink:href="#author"/>
    </roleSpec>
    <topicRef xlink:href="#shakespeare"/>
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#work"/>
    </roleSpec>
    <topicRef xlink:href="#hamlet"/>
  </member>
</association>
```

- Assoziation als Thema (Reification):

```
<topic id="will-wrote-hamlet-topic" >  
  <subjectidentity>  
    <subjectindicatorRef xlink:href=  
      "#will-wrote-hamlet" />  
  </subjectidentity>  
  <baseName>  
    <baseNameString>  
      Shakespeare's authorship of Hamlet  
    </baseNameString>  
  </baseName>  
</topic>
```
- will-wrote-hamlet-topic ist jetzt Topic, das weiterverwendet werden kann

- Occurrences:

```
<topic id="hamlet">
  <occurrence>
    <instanceOf>
      <topicRef xlink:href="#date-of-composition"/>
    </instanceOf>
    <resourceData>1600-01</resourceData>
  </occurrence>

  <occurrence id="hamlet-in-xml">
    <instanceOf>
      <topicRef xlink:href="#xml-version"/>
    </instanceOf>
    <resourceRef xlink:href=
      "http://www.csclub.uwaterloo.ca/u/relander/XML/hamlet.xml"/>
  </occurrence>
</topic>
```

- Merge von Topic Maps: `<mergeMap>`
`<mergeMap xl i nk: href=`
 `"http://www.shakespeare.org/plays.xtm">`
 `<topicRef xl i nk: href="#shakespeare" />`
 `<topicRef xl i nk: href="#drama" />`
`</mergeMap>`
- Verarbeitungsregeln im XTM Standard definiert