

kPeer (Knowledge Peers):

Informationssuche beim verteilten SemBloggen

Alexander Sigel, M.A. (sigel@wim.uni-koeln.de)

*Seminar für Wirtschaftsinformatik,
insbesondere Informationsmanagement*

Workshop „P2P Information Retrieval in Deutschland“

Leipzig, 21. November 2005

Allgemeiner P2P-Hintergrund des Seminars



siehe: Forschungsbereich Peer-to-Peer (P2P) Computing:
http://www.wim.uni-koeln.de/Peer-to-Peer_Computi.431.0.html



WIRTSCHAFTS
INFORMATIK

- AMCIS: Minitrack Economics of Peer-to-Peer Networks (2005-2006)
- HICSS: Minitrack Peer-to-Peer Computing (2003-2005)
- Lin, Fischbach & Shaw (2005): Peer-to-Peer Technologies for Business-to-Business Applications, in: Shaw, M.J. (ed.): Electronic Commerce and the Digital Economy, M.E. Sharpe, Amonk, NY, *forthcoming*
- Schmidt, Schoder & Fischbach (2005): Open Object Information Infrastructure: Wertschöpfung dank frei zugänglicher Objektinformationen
Einreichung zum Track "Betriebliche Anwendungen des P2P und Grid Computing" der MKWI 2006
- Schoder, Fischbach & Schmitt (2005): P2P Application Areas, in: Steinmetz & Wehrle (eds.): Peer-to-Peer Systems and Applications. Springer, 25-32
- Schoder, Fischbach & Schmitt (2005): Core Concepts in Peer-to-Peer (P2P) Networking, in: Subramanian & Goodman: P2P Computing: The Evolution of a Disruptive Technology, Idea Group Inc, Hershey, 1-27
- Schoder & Fischbach (2003): Peer-to-Peer Prospects, in: Communications of the ACM, 46(2) Februar 2003, 27-29
- Schoder & Fischbach (2003): Peer-to-Peer-Netzwerke für das Ressourcenmanagement, in: Wirtschaftsinformatik, 45(3), 313-323
- Schoder & Fischbach (2003): Peer-to-Peer, in: Wirtschaftsinformatik 44(6), 587-589
- Schoder, Schmitt & Fischbach (2003): Peer-to-Peer Netzwerke, in: WISU, 10/03, 1186 - 1191
- Schoder, Fischbach & Teichmann (Hrsg., 2002): Peer-to-Peer: Ökonomische, technische und juristische Perspektiven. Springer
- P2P-Hauptseminar (SS 2004)



Warum bin ich für unsere Gruppe auf diesem Workshop?

- Wir ...
 - möchten zwei interessante Anwendungsbereiche vorstellen, die **Informationssuche in P2P-Netzwerken** benötigen:
 - TedI: „Die föderative Bibliothek“
 - kPeer: „Verteiltes Sembloggen“
 - ... suchen Anregungen und Hinweise auf geeignete Ansätze und Werkzeuge für P2P und P2PIR, mit denen wir unsere Ideen dem Stand der Wissenschaft entsprechend umsetzen können
 - Bewertung der Eignung, ggf. Anpassung von **P2P-Infrastrukturen**
 - ... möchten mit anderen Gruppen mit ähnlichen Interessen ins Gespräch kommen und uns vernetzen
 - ... suchen die Zusammenarbeit
 - Entwicklung **semantischer P2P-Anwendungen auf Basis von Topic Maps**

Derzeit *keine* P2P-Drittmittelprojekte, aber drei *laufende* Dissertationen mit Bezug zu P2P

1. Dynamik von Peer-to-Peer Netzwerken (Kai Fischbach)

- Dynamisches Systemverhalten simulieren, verstehen und erklären

2. TedI (Technologien für ein dezentrales Informationsmanagement):

Ubiquitäre und P2P-basierte Technologien für ein dezentrales Informationsmanagement (Christian Schmitt)

- Ein Referenzszenario: „Die Föderative Bibliothek“

3. kPeer (Knowledge Peers):

Dezentrale Organisation, Integration und Aggregation von Wissen

mit semantischen Wissenstechnologien in verteilten, heterogenen Umgebungen (Alexander Sigel)

- Ein Anwendungsbereich: „Verteiltes SemBloggen“ (Semantic Blogging)

Was ist die gemeinsame Klammer von TedI und kPeer?

- Motivation:
Wertschöpfung durch neuartige und innovative Dienstleistungen
 - Veredelung, Verknüpfung und Integration von Daten, Informationen und Wissen
 - Aufgrund des emergenten Zusammenwirkens von Diensten und Nutzern werden deutliche Vorteile erwartet
- Ziel:
Dezentrales Informations- und Wissensmanagement (DKM)
 - überbetrieblich
 - autonome Akteure
- Mittel:
Anwendungsorientierter Einsatz semantischer P2P-Technologie
 - Möglichst Verwendung und Erweiterung bestehender Werkzeuge
 - RDF und Topic Maps
 - Wie treten hier *inhaltsbasierte Informationssuche-Prozesse in P2P-Netzen* auf?
 - voraussichtlich:
 - TMRAP zur Zusammenführung
 - tolog für die Anfragen

Was ist der Stand der beiden Vorhaben?

- Beide:
 - Konzeption, Szenarien- und Anforderungsdefinition: i.w. abgeschlossen
 - Auswahl von Infrastruktur-Werkzeugen zur Umsetzung: begonnen
 - Implementierung: bevorstehend
- TedI: Diplomarbeiten
 - RFID-Infrastruktur: abgeschlossen
 - Konzeption föderative Bibliothek: abgeschlossen
 - **Semantische P2P-Werkzeuge**: ausgeschrieben
- kPeer: Diplomarbeiten
 - Sembloggen mit Topic Maps (nicht verteilt)
Ziel: Zentralen Blojsom-Bloggingdienst um SemTags erweitern
 - **P2P-basiertes Sembloggen**: begonnen
Ziel: Anwendung auf verschiedenen Ebenen verteilen,
semantische Wissensdienste hinzufügen

TedI: Daten → Informationen

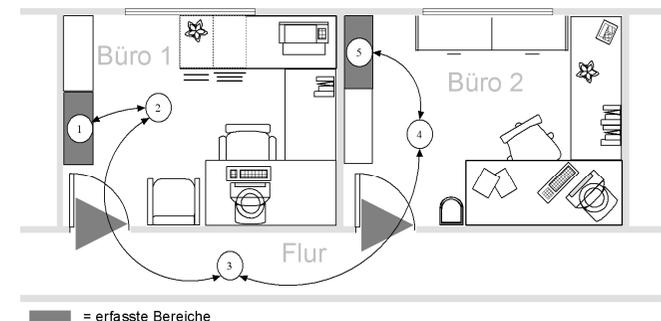
- **Motivation:** Die Datenflut in Datenschatten (aus Sensornetzwerken und von „smarten Objekten“ im Überschneidungsbereich von Ubiquitous/Pervasive Computing und Ambient Intelligence/Business) soll ökonomisch nutzbar gemacht werden
- **Ziel:** Erschließung externer Daten- und Informationsquellen für das dezentral organisierte, überbetriebliche Informationsmanagement
- Diese Objektinformationen
 - Informationen über physische Objekte, ihre Nutzungskontexte sowie semantische Annotationen von Nutzern des Systems über diese Objektesollen (verdichtet und semantisch angereichert) in einer noch aufzubauenden Infrastruktur (OOII - Open Object Information Infrastructure) als frei zugängliche Informationsbasis angeboten werden
- Die Objektinformationen sollen dezentral in P2P-Netzwerken abgelegt und Dritten zum Abruf und zur **Suche** angeboten werden
- Alle sollen diese Objektinformationen semantisch anreichern und integrieren können
- **Schwerpunkt:**
 - Daten- und Informationsknoten (als Infrastruktur „smarter“ Objekte)
 - dezentrale Datenhaltung in P2P-Netzwerken
 - „einfache“ Integration

<http://www.wim.uni-koeln.de/TeDi.334.0.html>
<http://www.wim.uni-koeln.de/uploads/media/TedI.pdf>

TedI: Referenzszenario Föderative Bibliothek

- Bewegungen von mit RFID-Tags versehenen Büchern an Lesegeräten vorbei erzeugen Sensordaten, die in Datenschatten gespeichert werden.
- Dezentrale Anwendungen sollen diese Daten auswerten und wiederum neue Informationen bereitstellen können
z.B.:
 - *Wo befindet sich ein Buch?*
 - *Wer hat es ausgeliehen?*
 - *Welche Bücher stehen bei einem Mitarbeiter nebeneinander im Regal? (potenzielle inhaltliche Ähnlichkeit)*
 - *Welche Personen arbeiten zu vergleichbaren Themen?*

http://www.wim.uni-koeln.de/uploads/media/Foederative_Bibliothek.pdf



kPeer: Informationen → Wissen

■ Motivation:

- Wenn mehrere Personen unabhängig voneinander Aussagen über geistige Gegenstände erstellen, über mehrere Quellen integrieren, veröffentlichen und suchen, so können die resultierenden semantischen Wissensdienste und deren Verknüpfung zu neuen Produkten und Dienstleistungen führen
- Die emergente Erstellung und die Exploration der Entwicklung vermaschter Wissensnetze für das dezentrale Wissensmanagement (DKM) in Wissensgemeinschaften steht erst am Anfang

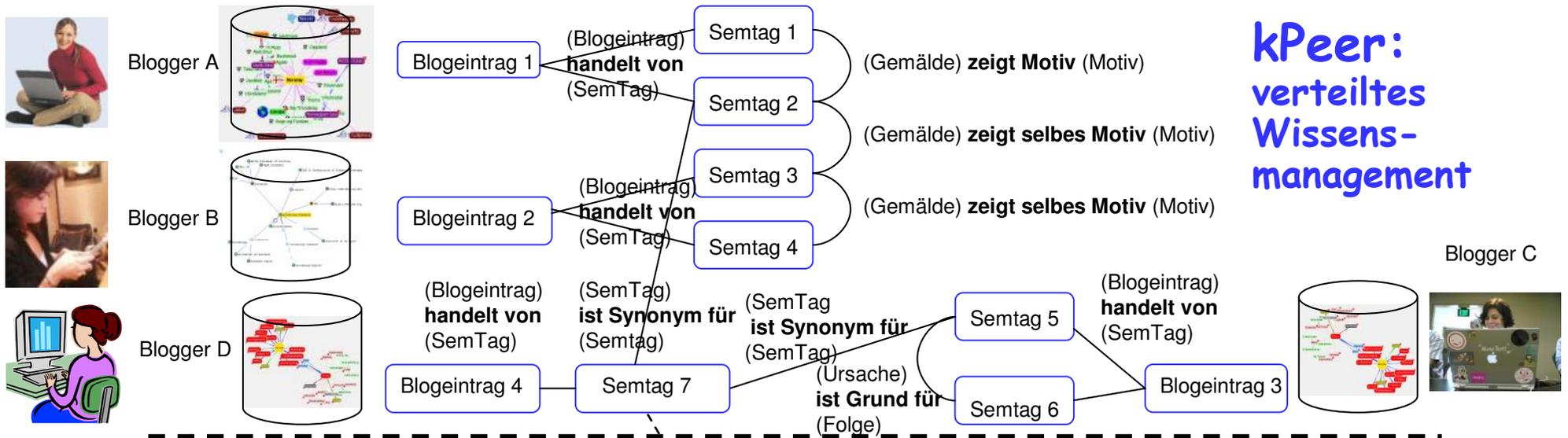
■ Ziel: Dezentral organisierte Wissensintegration und -aggregation aus unabhängigen, heterogenen Quellen mittels semantischer Wissenstechnologien zum Zwecke des DKM

- Unterstützung von Wissensgemeinschaften bei solchen Aktivitäten am Beispiel von Bloggen mit semantischen Auszeichnungen

■ Schwerpunkt:

- Wissensknoten
- Semantische Annotationen menschlicher Akteure
- Semantische Integration

TedI und kPeer: mögliches Zusammenspiel



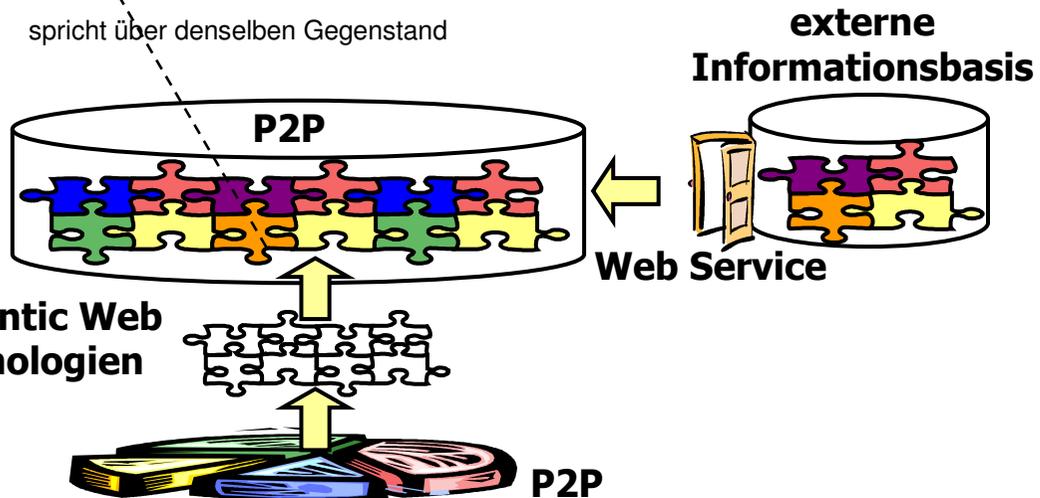
kPeer:
verteiltes
Wissens-
management

TedI:
verteiltes
Informations-
management

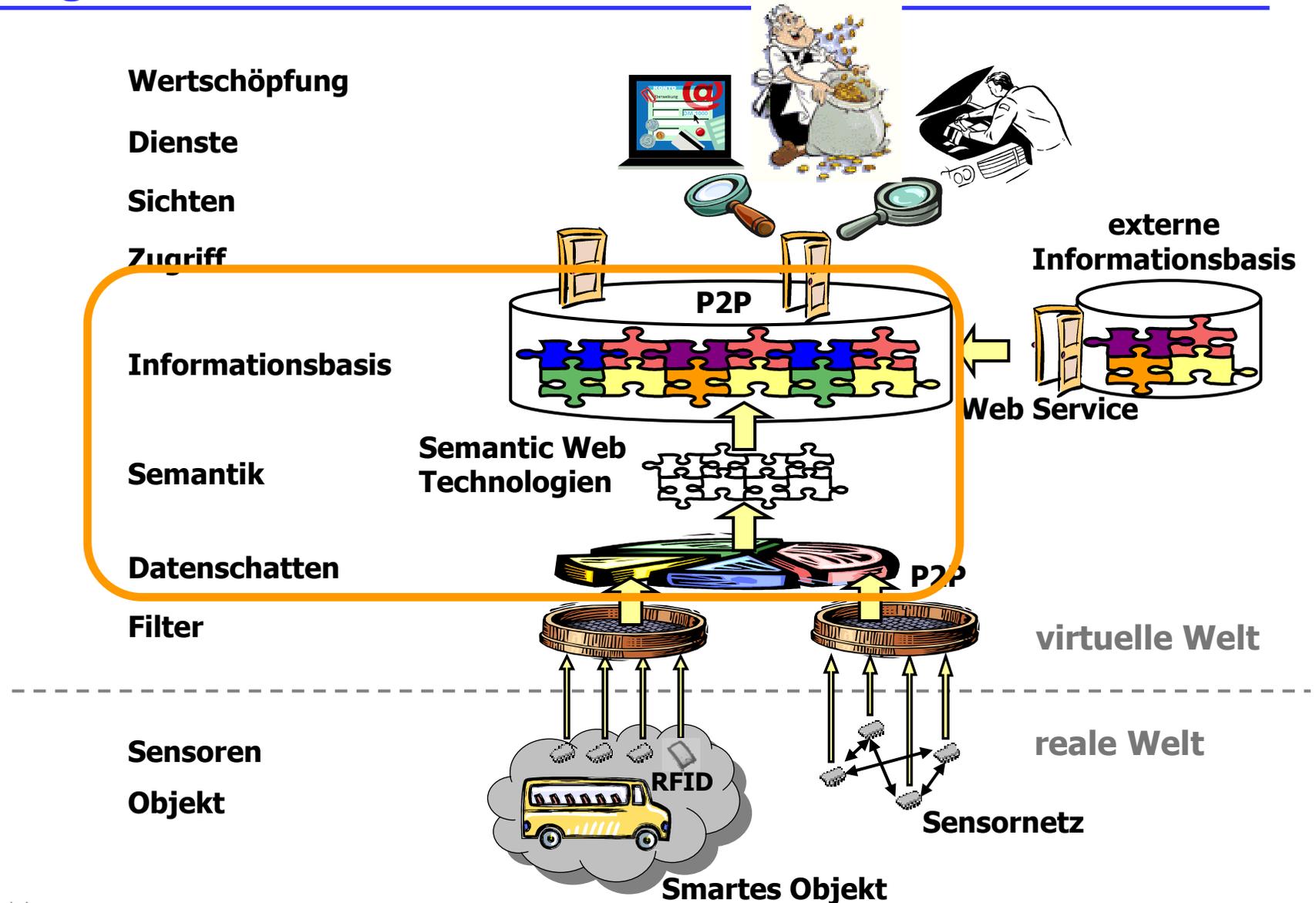
Informationsbasis

Semantik

Datenschatten



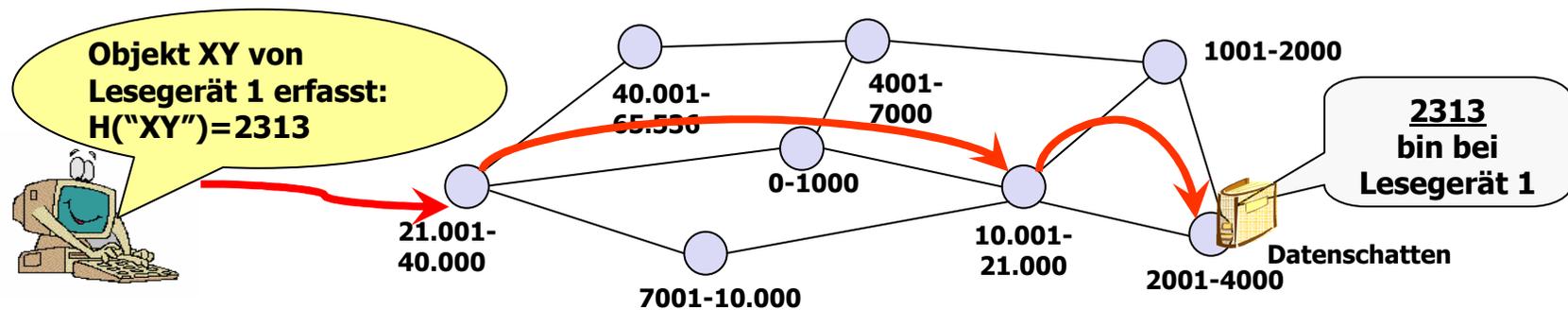
TedI: Wo gibt es hier inhaltsbasierte Suche in P2P-Netzen?



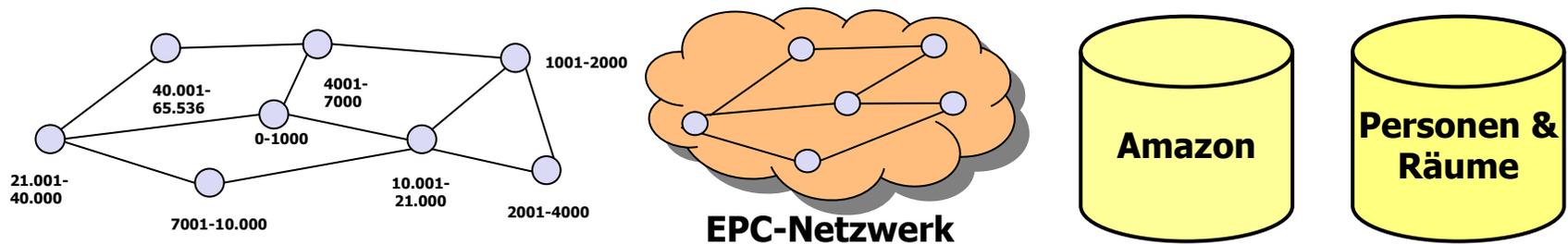
TedI: Speicherung und Zugriff in einem strukturierten P2P-Netzwerk

- Strukturiertes P2P-Netzwerk

- Dezentrale Speicherung dieser Informationen



- Freier Zugriff auf Informationen



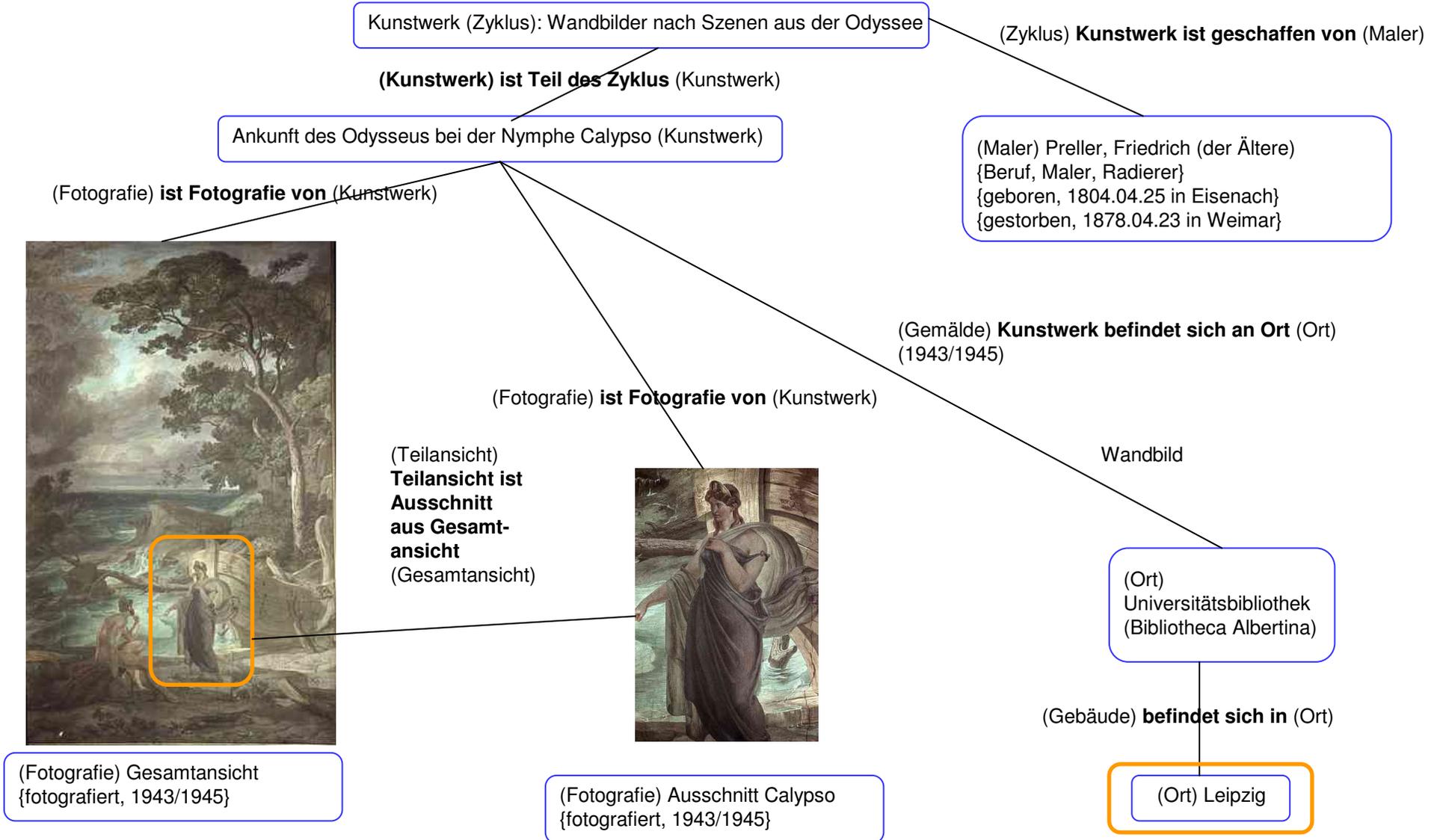
kPeer: Anforderungen

- Anstatt einen eigenen Blogeintrag mit einem einfachen Tag zu kennzeichnen, sollen alle zu eigenen, aber auch zu fremden Blogeinträgen semantisch ausgezeichnete Aussagen über geistige Gegenstände, die Elemente in einfachen Ontologien referenzieren, öffentlich ablegen, weiterverarbeiten, zusammenführen/verbinden und **suchen** können
 - Semantik statt Tags
 - Soziale Komponente
- Anstatt Einträge aus zentral definierten Ontologien zu verwenden, sollen alle:
 - einfache Ontologien anlegen können, die von ihnen definierten Elemente der Ontologien in Bezug setzen können zu Elementen in Ontologien Dritter, sowie darin **suchen** können
 - Folksonomy
 - → semantische Interoperabilität

Beispiel: Eine Identitätserklärung, dass ein geistiger Gegenstand, der von einem anderen Peer veröffentlicht wurde, dem eigenen entspricht.

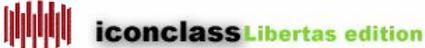
SemBlogger A: sembloggt über ein zerstörtes Kunstwerk ...

(<http://zi.fotothek.org>)



SemBlogger A: ... und fügt ICONCLASS 94I21 hinzu

http://iconclass.bureau.knaw.nl/libertas/ic?task=getnotation&datum=94I211&style=notationbb_hab.xsl&taal=de






©KNAW

Notation:

Keyword:

Context of: 94I21

9	Klassische Mythologie und Antike Geschichte
94	die griechischen Heldensagen (I)
94I	die Abenteuer der Griechen nach dem Fall Trojas: die Irrfahrten des Odysseus (Homer, Odyssee)
94I2	die Irrfahrten des Odysseus: von Calypso bis zu den Phaiaken
94I21	auf der Insel Ogygia wird Odysseus von der Nymphe Calypso freundlich empfangen (Homer, Odyssee XII)

Systematic References

95A(TELEMACHUS)31	Telemachus und Calypso
-------------------	------------------------

Suche nach Illustrationen zur Notation **94I21** in:
[Festkultur Online - HAB Wolfenbüttel](#)




Quick Links to Main Categories:
[\[0\]](#) [\[1\]](#) [\[2\]](#) [\[3\]](#) [\[4\]](#) [\[5\]](#) [\[6\]](#) [\[7\]](#) [\[8\]](#) [\[9\]](#)

Children

94I21(+)	KEY
94I211	Merkur befiehlt Calypso, Odysseus freizulassen (Homer, Odyssee V)
94I212	Odysseus baut ein Floß

Keywords

- antike Geschichte
- klassische Antike
- Geschichte
- Mythologie
- Legende
- griechische Legende
- Held
- heroische Legende
- Zyklus
- episch
- Homer, Odyssee
- wandern
- Odysseus
- segeln
- zurueckkehren
- Homer, Odyssee 12
- empfangen (jemanden)
- Calypso
- Insel
- Ogygia (Insel)

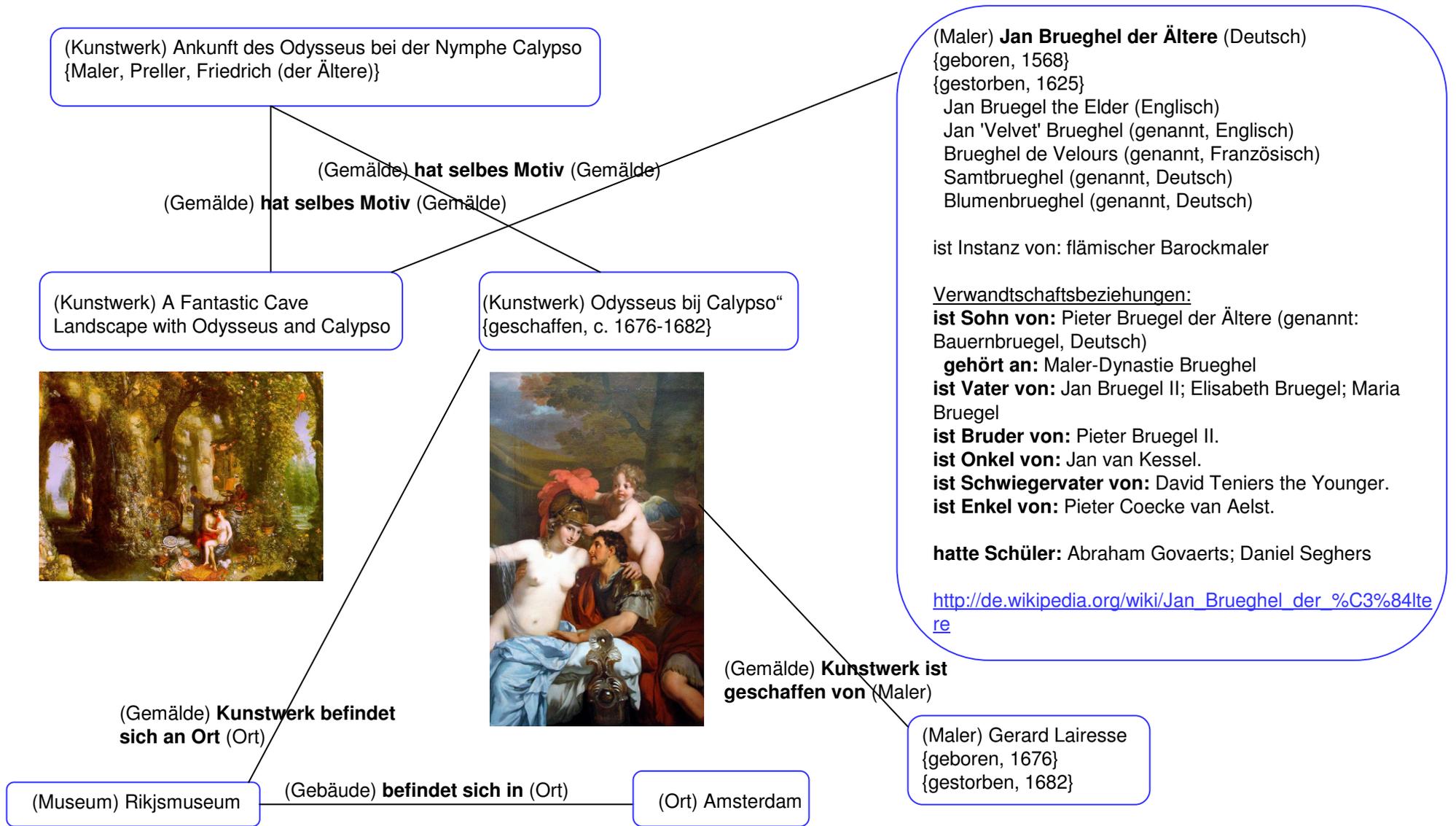
ICONCLASS 94I21

(Motiv) **wird dargestellt in** (Gemälde)

Ankunft des Odysseus bei der Nymphe Calypso (Kunstwerk)



SemBlogger B: sembloggt, dass zwei weitere Bilder dasselbe Motiv haben



SemBlogger C: sembloggt über das dasselbe Motiv

- Sie fügt ihrem textuellen Posting <http://stellier.blogspot.com/2005/10/odysseus-and-calypso.html> folgende Semantik hinzu:

Calypso
(Nympe in der griechischen Mythologie)

(Ursache) **ist Grund für** (Folge)

7 Jahre Verzögerung der Rückkehr des Odysseus



<http://photos1.blogger.com/blogger/5711/1620/1600/calypso%20odysseus1.jpg>

SemBlogger D: ...

- definiert ihr Semtag

Kalypso (Mythologie)

PSI: psi.semblogging.de/blogger_d/1481...

Wikipedia: http://en.wikipedia.org/wiki/Kalypso_%28mythology%29

ICONCLASS 94I21:

http://icontest.bureau.knaw.nl/libertas/ic?task=getnotation&datum=94I21&style=notationbb_hab.xsl&taal=de

als in ihrer Sicht identisch mit dem SemTag von Blogger C.

Kalypso
(Mythologie)

(SemTag) ist **Synonym** für (SemTag)

Calypso
(Nympe in der griechischen Mythologie)



Beispiele für Suchanfragen in kPeer, die sich auf verteiltes Wissen beziehen

- Welche SemTags sind über alle SemBlogger als Unterbegriffe der Irrfahrt des Odyseus bekannt?
 - *Das ICONCLASS-Motiv 94I21 (ist Teil von)*
- Was wissen alle Peers zusammen über: Calypso?
 - *Virtuelle Integration aller Aussagen über einen geistigen Gegenstand, auch wenn sie aus mehreren semantisch heterogenen Quellen stammen und unabhängig voneinander erstellt wurden*
- Welche Blogs sprechen über das ICONCLASS-Motiv 94I21?
 - *Nur durch die Identitätserklärung von Semblogger D ist der Blogeintrag von Semblogger C mit ICONCLASS ausgezeichnet*
- (fortgeschritten:)
Welche Beziehung besteht zwischen Daniel Seghers und Gerard Lairesse?
 - *Laut Wikipedia war sein Lehrer Jan Brueghel der Ältere, der laut Semblogger B ein Bild mit demselben Motiv, nämlich ICONCLASS 94I21 gemalt hat*

Geplante technische Realisierung in TedI

- Datenhaltung
 - voraussichtlich in RDF
- Praktisch nutzbare Werkzeuge für Semantic Overlay Networks (SONs) gesucht
 - vgl. Aberer & Cudré-Maroux: Tutorial bei VLDB
<http://www.vldb2005.org/program/slides/thu/s1367-aberer.pdf>
<http://lsirwww.epfl.ch/SON/>
 - Query Evaluation
 - RDFPeers
 - PIER
 - Edutella
 - Semantic Mediation in SONs (PDMSs - Peer Data Management Systems)
 - PeerDB
 - Hyperion
 - Piazza
 - Semantic Gossiping (Chatty Web)
 - GridVine
- weitere z.B.:
 - DBin
 - InfoQuilt
 - P-Grid
 - SWAP (Bibster)
 - Semantic Desktop-Ansätze
- Möglichst mit Unterstützung für
 - P2P-Semantic Web-Services

Technische Überlegungen in kPeer

- Bisher (zentral):
 - Wir betreiben einen zentralen SemBlog-Dienst (Blojsom + TMAPI + TM4J)
 - Das Wissen eines jeden Peers ist jeweils in seiner lokalen Topic Map gespeichert
 - Es gibt ein zentrales PSI-Verzeichnis
 - Wir betreiben einen zentralen Aggregations-Dienst:
Der beste Stand des Gesamtwissens wird nach Bedarf (periodisch und bei Events) auf Basis der Antworten aller Peers in eine zentrale Topic Map zusammengeführt
 - Alle können diese zentrale Topic Map abfragen und dabei verschiedene Sichten einnehmen
 - Alle Abfragen und Inferenzen erfolgen in tolog auf der zusammengeführten Topic Map
- Angedachte weitere P2P-Verteilungsmöglichkeiten:
 - Alle können einen Semblog-Dienst betreiben, d.h. die Topic Maps der Peers liegen nicht mehr beim selben Dienstprovider
 - Mehrere autonome Aggregatoren können Teilstände zusammenführen und dabei auch hierarchisch oder anderweitig zusammenarbeiten
 - Alle können alle Anbieter (mit tolog) abfragen
 - Alle können PSI-Verzeichnisse anbieten

kPeer: Anwendung verteilter Topic Maps und P2P

- SemBlogger ~ kPeer
- Peers tauschen Topic Map-Fragmente über TMRAP - Topic Maps Remote Access Protocol
- Bisläng:
 - Ein zentraler Dienst, der die Semantik der Topic Maps versteht
- Motivation für einen stärker verteilten Ansatz?
 - DKM: Autonome Akteure arbeiten verteilt und wünschen ihre Informationshoheit (z.B. Abteilung möchte eigene Ontologiesicht)
 - Voraussichtlich demnächst:
Mehrere Dienste, die jeweils die Semantik der Topic Maps verstehen
 - Replikation, damit einmal ausgezeichnete Aussagen nicht verloren gehen
- Noch nicht gut verstanden:
 - Wie müssen verteilte Topic Map P2P-Netzwerke aussehen, damit auch dezentrale Informationssucheprozesse effizient unterstützt werden können?
 - „Semantic Overlay für Topic Maps?“ oder Topic Maps als Black Box für P2P?