

What did you call it again?

Language Use in Group Information Management

Emilee Rader
Pre-Proposal Meeting
October 3, 2007

“...Because there is a group of us using the site I find it difficult to keep up with what is located where, and so ask people to email me direct copies of materials I need...” (CTools project site user)

the problem

- People who work together often need to share files; one way to do this is via storage space on a server
- Organizations are adopting user-contributed content systems
 - enterprise content repositories
 - software code repositories
 - wikis and blogs used as “group memory”
- These systems are relied upon by group members for keeping their information safe and making it available when it is needed
- Users can spend a significant amount of time searching for or re-creating a single missing document

so what?

- Information producer and consumer **roles**, not always the same person
- Sharing as making content **available** to multiple potential 'recipients', not sending
- Many contributions are to some extent **discretionary**
- Unstructured or **uncontrolled** organization scheme and vocabulary
- **Natural language** (names, labels, descriptions) as identifier

- File naming conventions agreed upon by a group in principle are difficult to adhere to in practice
- There's lots of variability in language use; people don't often agree on names for the same things
- Common ground allows people to negotiate meaning in conversation
 - shared immediate context
 - shared past experience
 - community or category membership
- People label things differently when directed to do so for themselves, vs. some unknown future person

literature

- Different people fill information producer and consumer roles at different times
- This means there are multiple potential recipients, all with unique knowledge, context, and needs
- ‘Packaging’ for reuse by others is more difficult when there can be multiple possible information consumers
- Each producer is free to ‘package’ however they feel like it and tends to do so according to their own style and preferences

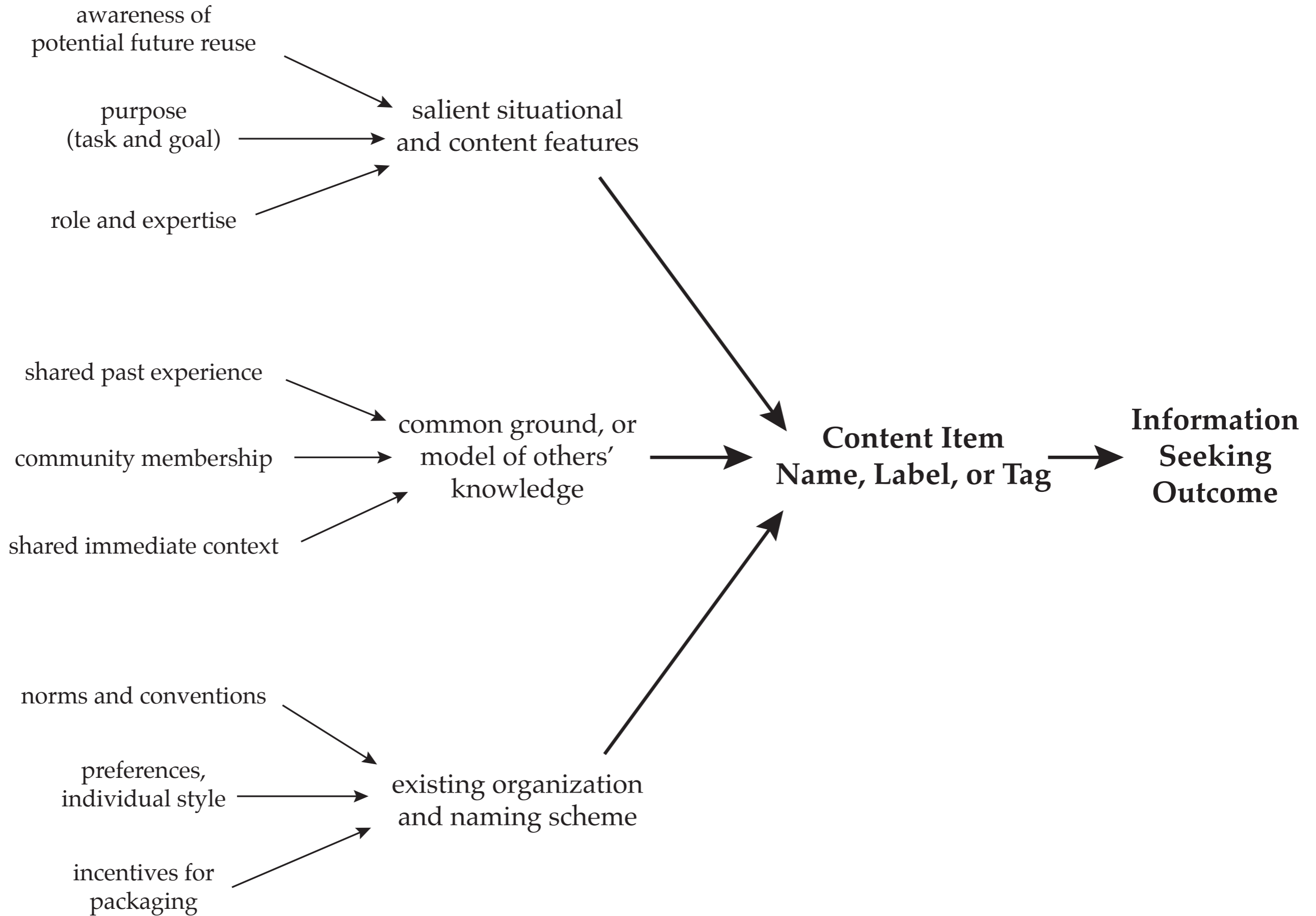
- Groups of users who are known to each other, different from PIM. Not much research on this yet
- Also different from information seeking on the web and in libraries and archives
- Discretionary contributions: similar to other CSCW systems
- Influences that shape language use have not been studied in an information management context
- Knowledge management research tends to identify factors and outcomes at the organization level, rather than the workgroup

- Language supplies a “handle” or identifier for content items
- Language represents salient aspects of items’ content or context
- Language manifests relationships among content items, and conveys information about the values of the group
- Language connects information producers and consumers, via the repository

why language?

The purpose of this research is to improve users' ability to find information in situations where multiple users access shared information resources via a user-contributed content system, by investigating how language use affects the organization, structure and seeking of information.

purpose



How does language use affect information sharing in user-contributed content repositories?

1. How do groups organize and share information via an online repository of user-contributed content? What problems do individual members encounter when seeking information within the repository?
2. How do the type of common ground and the intended audience affect file and folder name choices, and the effectiveness of those names for information seeking?

questions

Study 1 (field study): How do groups organize and share information via an online repository of user-contributed content? What problems do individual members encounter during information seeking within the repository?

Study 2 (experiment): How do the type of common ground (from community membership or shared past experience) and the intended audience affect file and folder name choices?

Study 3 (experiment): How does the influence of common ground and intended audience on file and folder names affect information seeking outcomes in a shared online repository?

awareness of potential future reuse

purpose (task and goal)

role and expertise

salient situational and content features

shared past experience

community membership

shared immediate context

common ground, or model of others' knowledge

Content Item
Name, Label, or Tag

Information Seeking Outcome

norms and conventions

preferences, individual style

incentives for packaging

existing organization and naming scheme

Study I: Field

- How do groups organize and share information via an online repository of user-contributed content? What problems do individual members encounter during information seeking within the repository?
- Participants: users of CTools project sites
- Sources of Data: interviews, search tasks, event log analysis
 - frequency and recency of use
 - ownership of the item
 - depth in the hierarchy
 - descriptiveness of filename

study 1

b-school club

in its 2nd year

9 members

43% days with an event

no “manager”

total items: 331

total events: 1516

‘new’ events: 24%

‘read’ events: 66%

‘delete’ events: 5%

biomed engr. lab

in its 4th year

~20 members

75% days with an event

lab manager

total items: 395

total events: 2416

‘new’ events: 25%

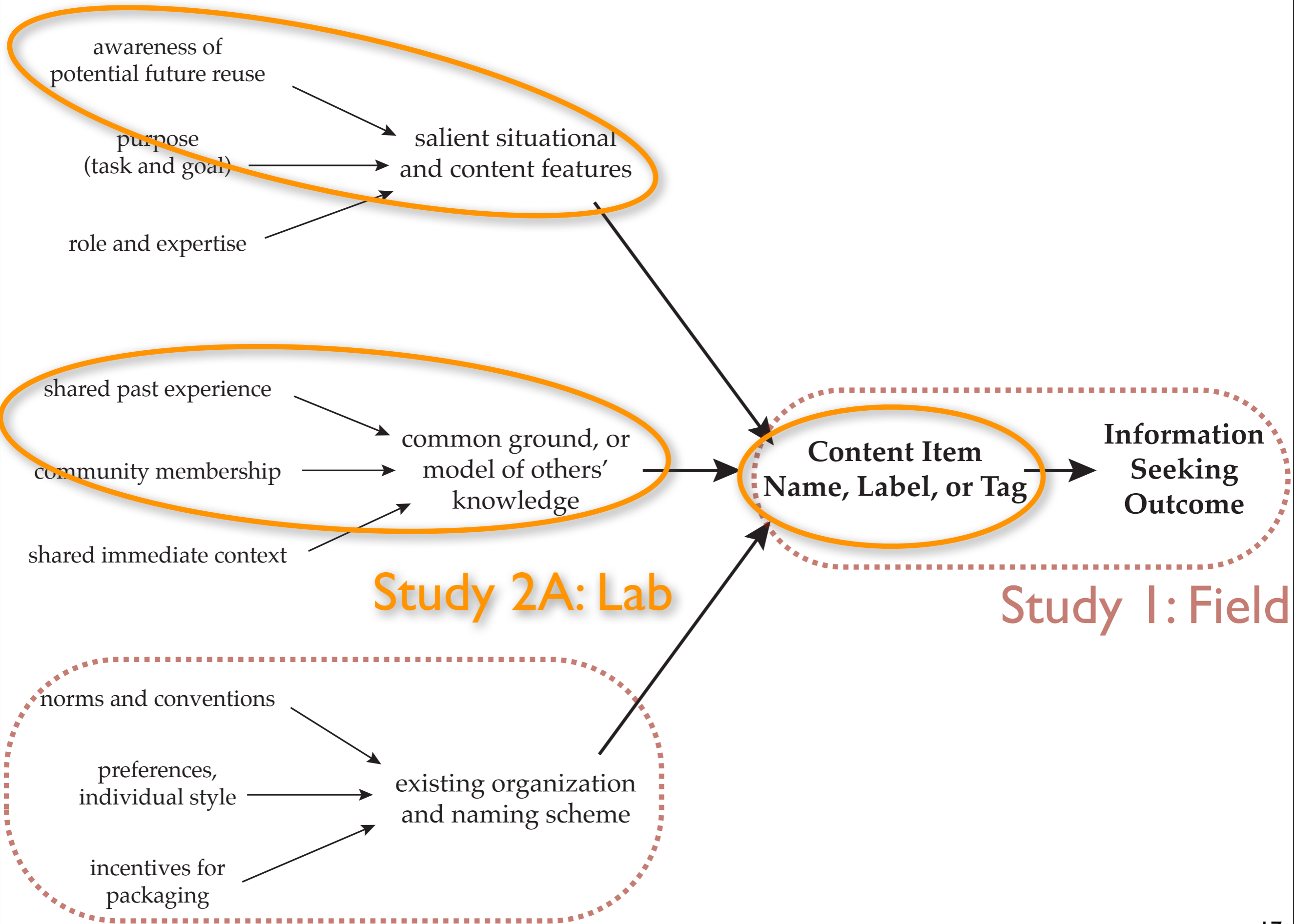
‘read’ events: 58%

delete events: 4%

descriptives

- **Measures and Analysis**
 - Search task think-aloud protocols
 - Quantitative analysis of success/failure
 - Qualitative coding of interview transcripts
 - Event log analysis
- **Expected Outcome**
 - Assessment of how different file characteristics relate to the “finding” problem (or not!)
 - Taxonomy of types of information management tasks, problems and issues

study 1

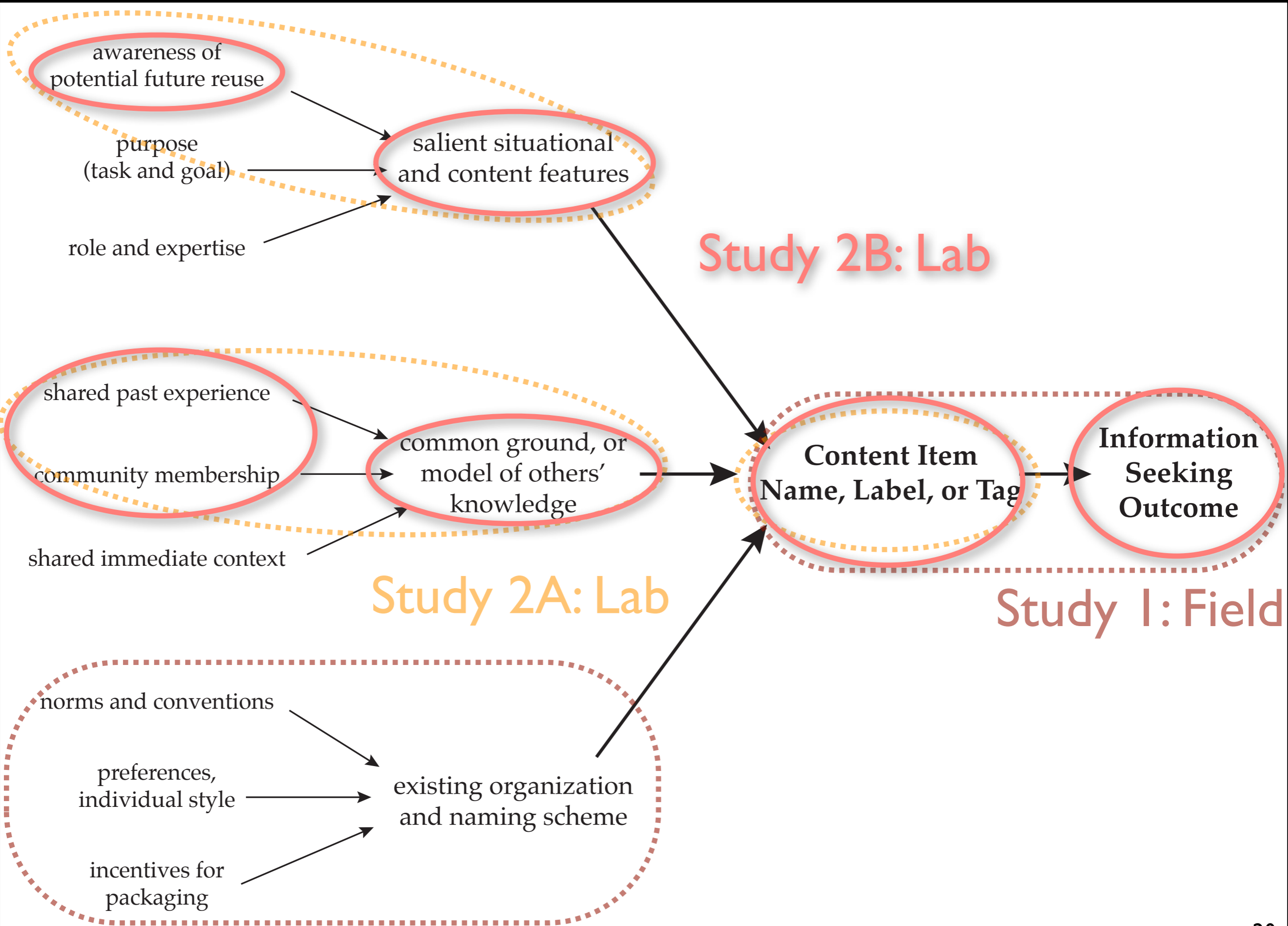


- How do the type of common ground (from community membership or shared past experience) and the intended audience affect file and folder name choices?
- Method
 - name documents and organize them into folders
 - post-experiment questionnaire assessing “model of others’ knowledge”
- Measures and Analysis
 - timing data will be automatically captured
 - qualitative content analysis of the names and folder labels
 - quantitative distance measure to compare hierarchy structures
- Expected Outcome

3 (cohort) x 5 (audience) between subjects design

	No Audience	For Self	For Your Class	For Last Year's Class	For General Public	(total)
MSI Class of 2008	control 10	control 10	SPE 10	CM 10	NCG 10	(50)
MSI Class of 2009	control 10	control 10	SPE 10	CM 10	NCG 10	(50)
Non-SI Grad Student	control 10	control 10	X (no subjects)	X (no subjects)	NCG 10	(22)
(total)	(30)	(30)	(20)	(20)	(30)	(130)

SPE = Shared Past Experience Common Ground
 CM = Community Membership Common Ground
 NCG = No Common Ground



- How does the influence of common ground and intended audience on file and folder names affect information seeking outcomes in a shared online repository?
- Method: participants from Study 2A will return 4-6 weeks later, and complete search tasks using hierarchies created in the previous experiment
 - created with no audience in mind (baseline)
 - created for oneself (baseline)
 - created for oneself by someone else (baseline)
 - Shared past experience: created by someone in the same cohort
 - Community membership: created by someone in other SI cohort
 - No common ground: created by someone outside SI
- Timing data will automatically be captured
- Expected Outcome

3 (cohort, between) x 6 (audience, within) mixed design

and, created for self by someone else

	No Audience	For Self	For Your Class	For Last Year's Class	For General Public	(total)
MSI Class of 2008	control 10	control 10	SPE 10	CM 10	NCG 10	(50)
MSI Class of 2009	control 10	control 10	SPE 10	CM 10	NCG 10	(50)
Non-SI Grad Student	control 10	control 10	X (no subjects)	X (no subjects)	NCG 10	(22)
(total)	(30)	(30)	(20)	(20)	(30)	(130)

SPE = Shared Past Experience Common Ground
 CM = Community Membership Common Ground
 NCG = No Common Ground

- Adds to our understanding of language use in situations not traditionally thought of as ‘communication’
- Explores factors that affect the usage and usefulness of a growing category of info management systems
- Combines psychology and LIS perspective to investigate ‘unstructured’ organization schemes
- Results will be used to inform system architecture and interface design choices (future work)

contribution

completion	activity
12/2007	IRB approval, pilot procedures
1/2008	Participant recruiting
4/2008	Data collection
8/2008	Data analysis
12/2008	Writing
2/2009	Finished!

timeline

Morae Recorder Software	\$195
Subject Payments, Study 1 8-10 different CTools project sites, 1 informant (\$25) and 4 users per site (\$15); plus pilot sessions	\$900
Subject Payments, Study 2A 100 participants (\$10 ea.), plus pilot sessions	\$1200
Subject Payments, Study 2B 60 participants @ \$10 ea., plus pilot sessions	\$700
TOTAL	~\$3000

Thank You!