An ontological approach to information access control and provenance

Bill Andersen Fabian Neuhaus

October 22, 2009

OIC 2008 . . .

What can the IC learn from the success of ontologies in the biomedical domain?

. . . not much

One reason:
biological ontologies deal with open and epistemically stable information

OIC 2008 . . .

What can the IC learn from the success of ontologies in the biomedical domain?

. . . not much

One reason:
biological ontologies deal with open and epistemically stable information

OIC 2008 . . .

What can the IC learn from the success of ontologies in the biomedical domain?

. . . not much

One reason:
biological ontologies deal with open and epistemically stable information

IC information is not open

The Information Dilemma

Responsibility to share information (ICD 501)
vs.
Responsibility to protect information
(incl. methods & sources)

Need for access control

IC Information is not Epistemically Stable

- information provided by unreliable sources
- logically inconsistent information

Epistemically Stability

- Biologists: What is the case?
- IC: Who said what?

Need for provenance tracking

Regards from Bill

Goal

Ontology language that supports

- multi-level access control
- provenance tracking
- reasoning with logically inconsistent information

Features for access control and provenance tracking are fully integrated in the language, not a separate meta-level layer.

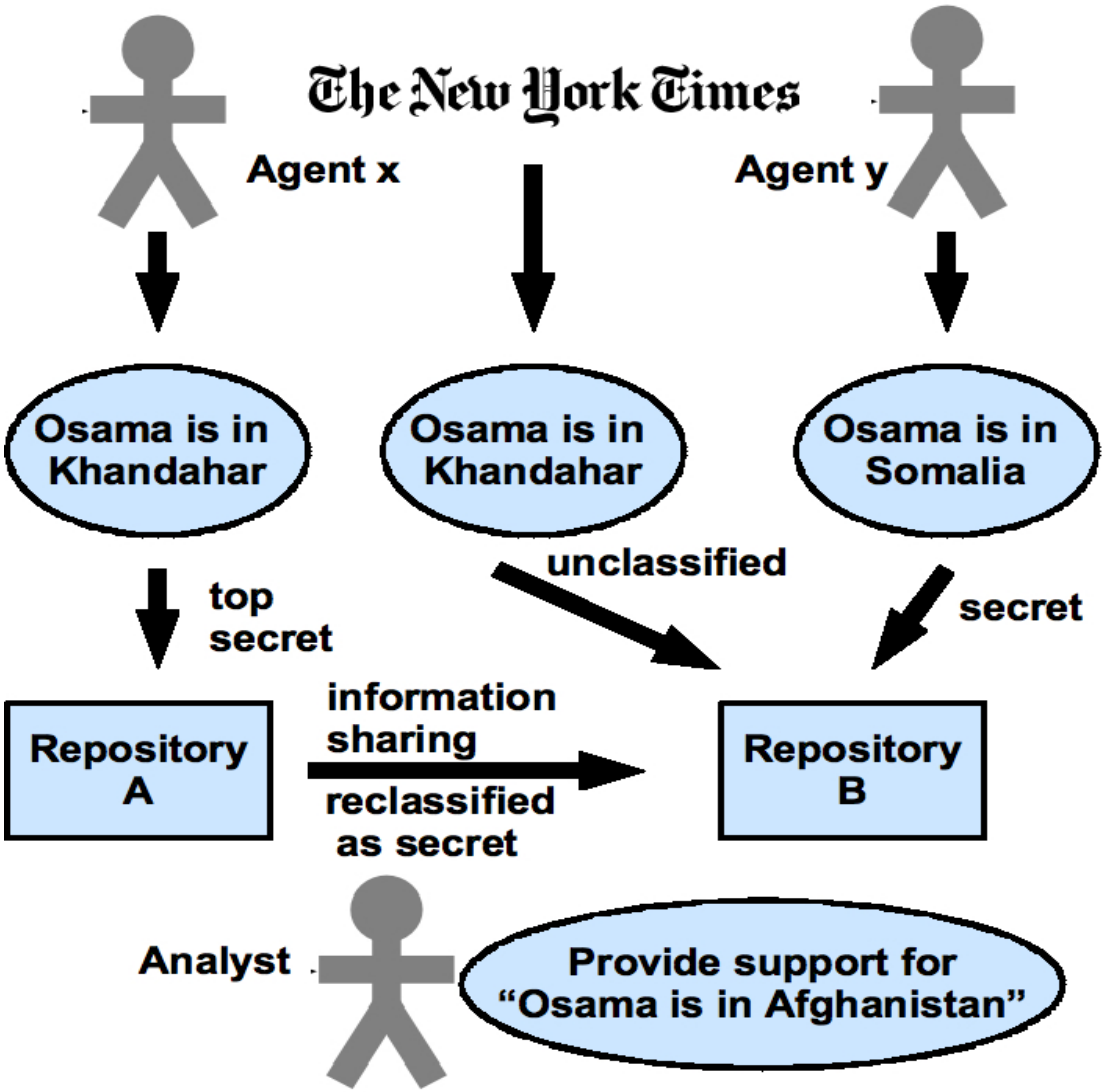
Introduction

Example

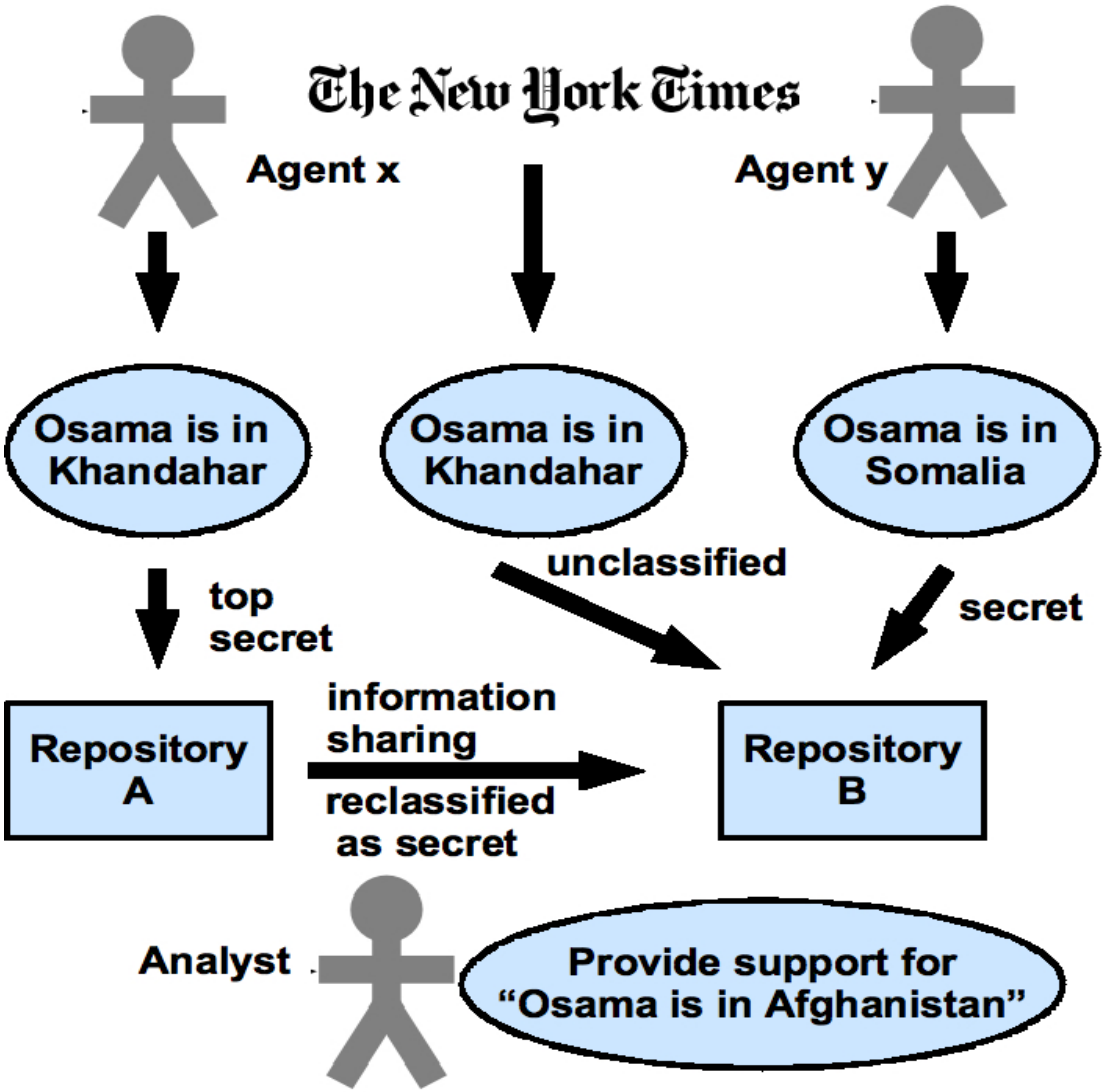
Ontological considerations

Formalization

Running Example

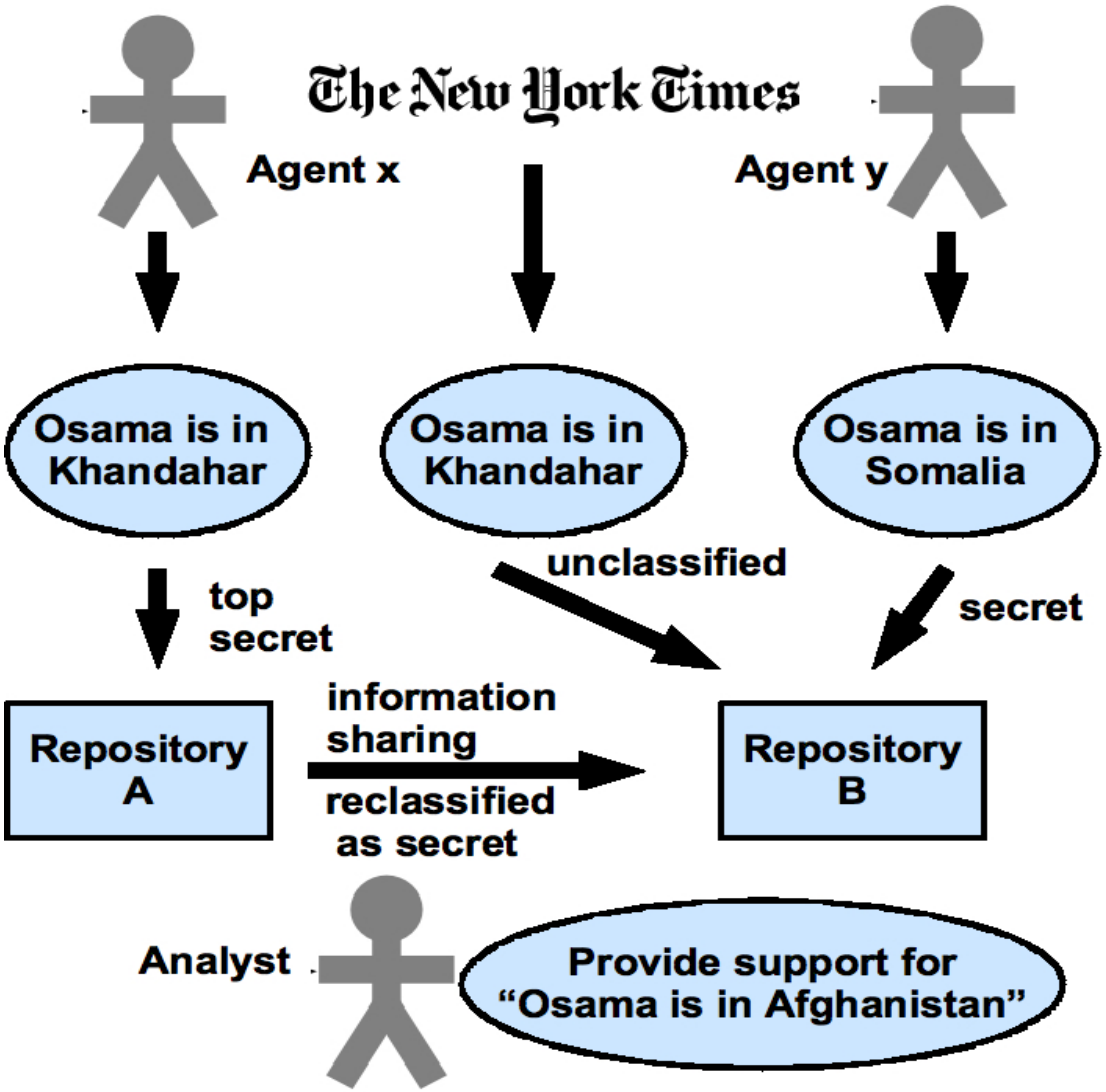


Running Example



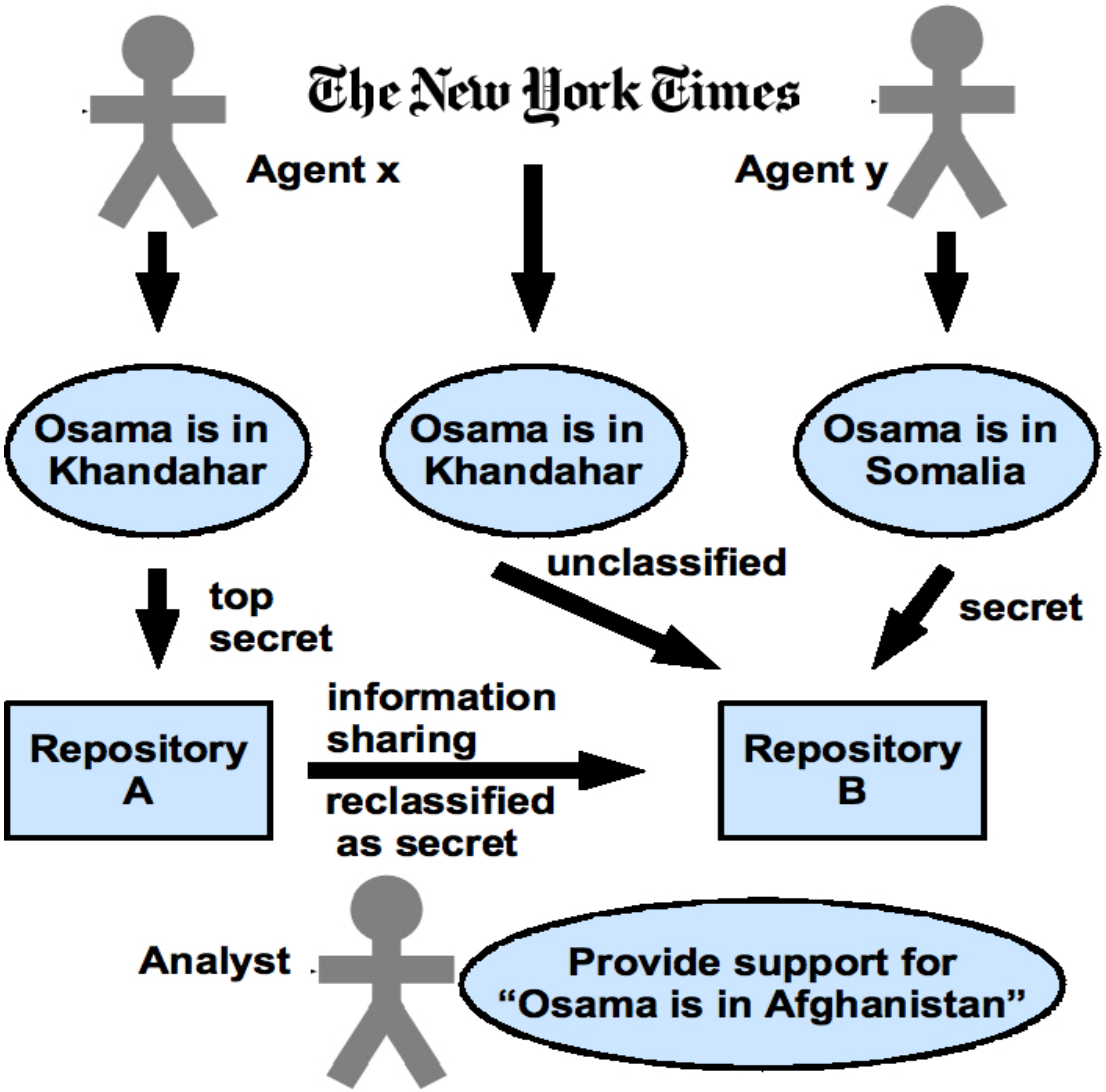
Challenge 1: Reasoning with geospatial information

Running Example



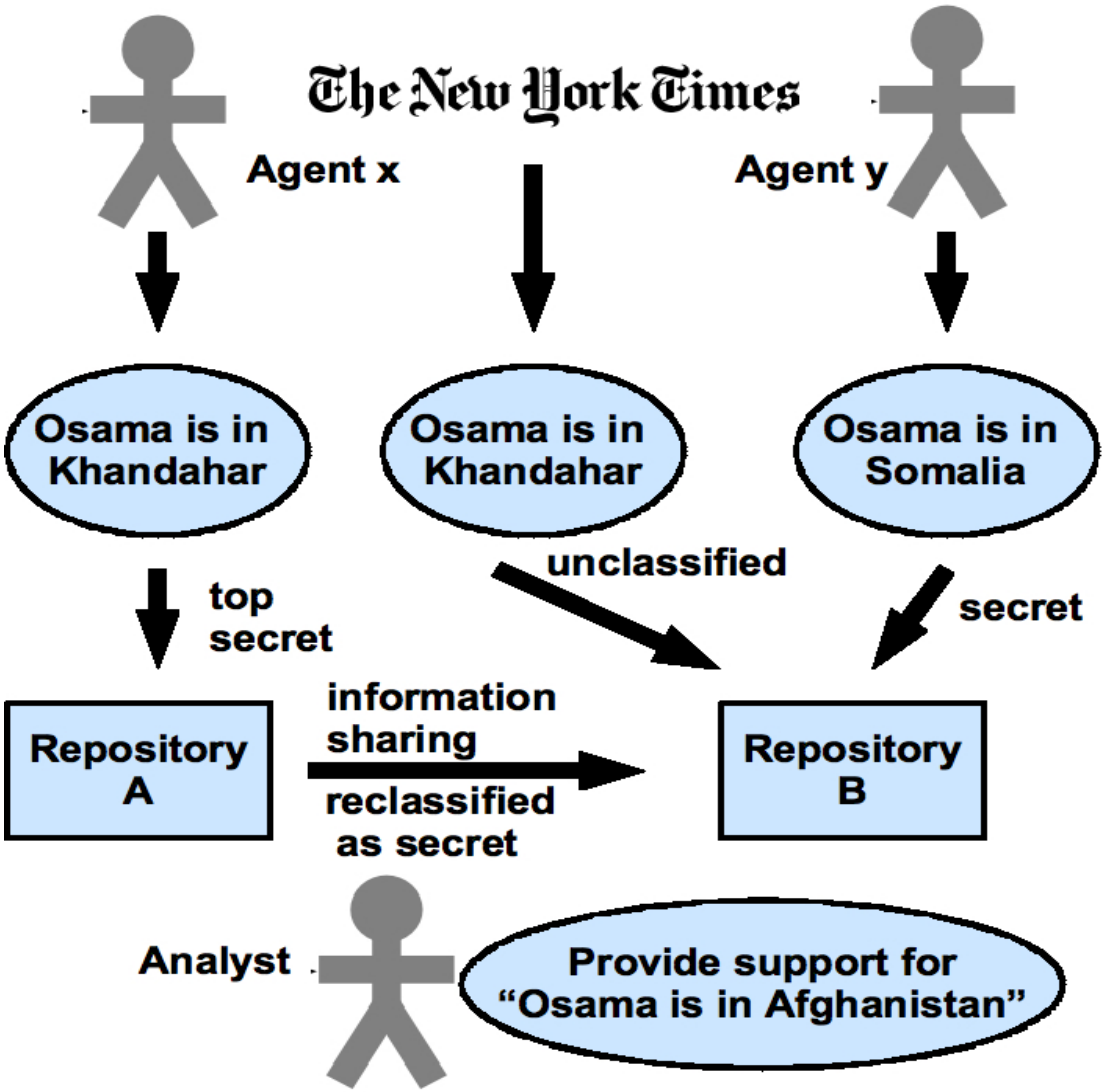
Challenge 2: Reasoning with inconsistent information

Running Example



Challenge 3: Security classification of analyst impacts query

Running Example



Challenge 4: One record in B is a 'copy' of the record in A

Introduction

Example

Ontological considerations

Formalization

What is Information

Information is abstract

- no spatio-temporal location
- no causal powers
- other examples: the natural numbers, geometric shapes, moral values

Ontology of Access Control II

It is impossible to control access to information!

Objects of access control are physical entities that encode information; e.g. paper documents, arrangements of electrons in a chip etc.

Ontology of Access Control II

It is impossible to control access to information!

Objects of access control are physical entities that encode information; e.g. paper documents, arrangements of electrons in a chip etc.

Proposition, Sentence Type, Sentence Token

proposition: a unit of information that is true or false

sentence type: a well-formed expression of a language that can be asserted by speakers to express a proposition

sentence token: a physical entity (or an arrangement of physical entities) that instantiates a sentence type

Example: How many sentences do you count?

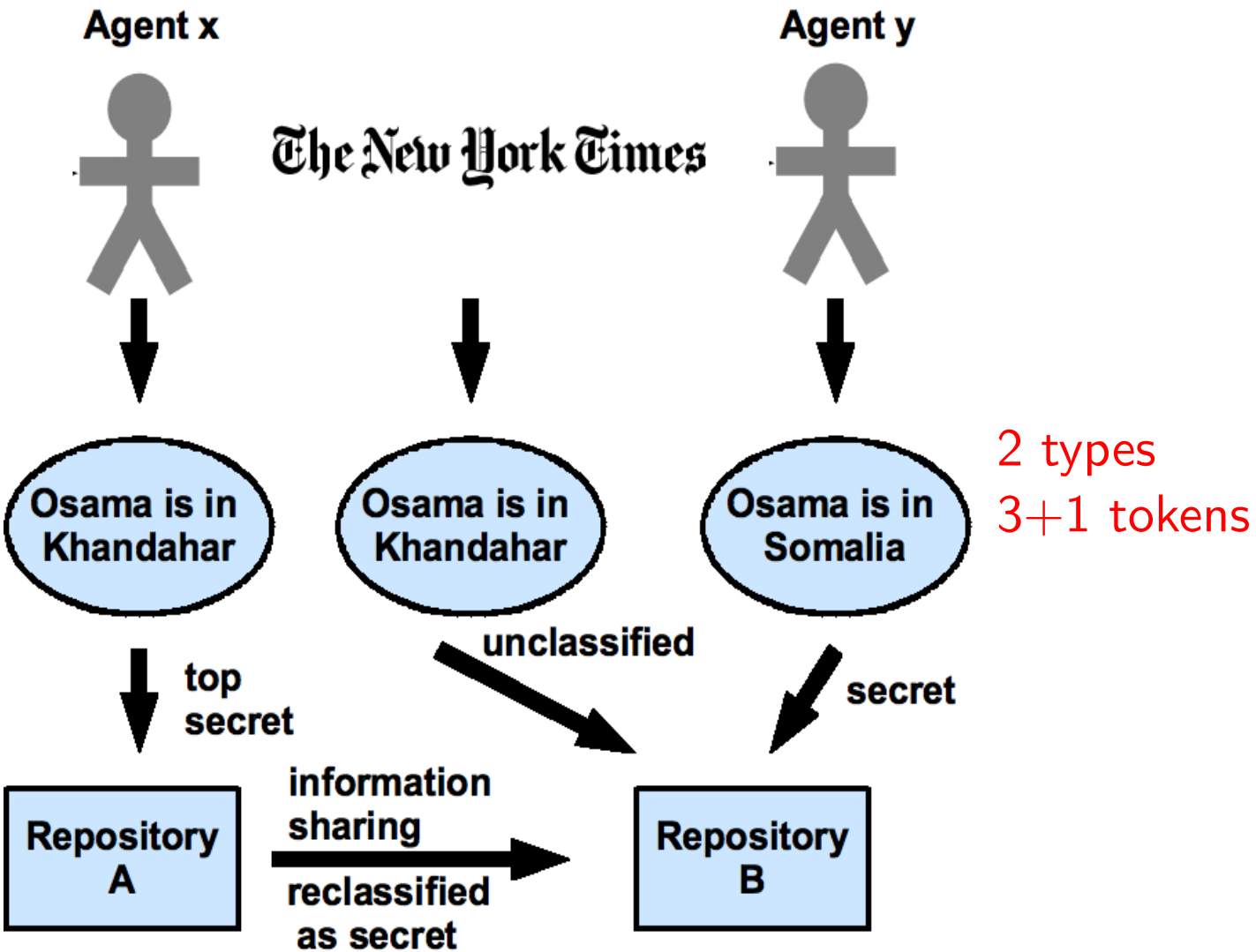
The cat is on the mat.

Die Katze ist auf der Matte.

The cat is on the mat.

Sentence tokens are the primary bearers of security properties and access control!

Running Example



What you can do with tokens . . .

- Copying
- Recoding (e.g., English to German, JPG to PDF)
- "Write down" (e.g., blackening sentences)
- Syntheses (e.g., summary, **automatic reasoning**)

Reformulation of the Problem

Provenance: causal history of tokens

Access control:

- set of security classifications of tokens (levels, compartments)
- security classification impact the processing of tokens
- inheritance of security classifications along causal history

Automatic Logical Reasoning

Osama is in Khandahar <top secret>
IF somebody is in Khandahar THEN he is in Afghanistan

New token:
Osama is in Afghanistan <???>

Automatic Logical Reasoning

Osama is in Khandahar <top secret>
IF somebody is in Khandahar THEN he is in Afghanistan

New token:
Osama is in Afghanistan <???>

Introduction

Example

Ontological considerations

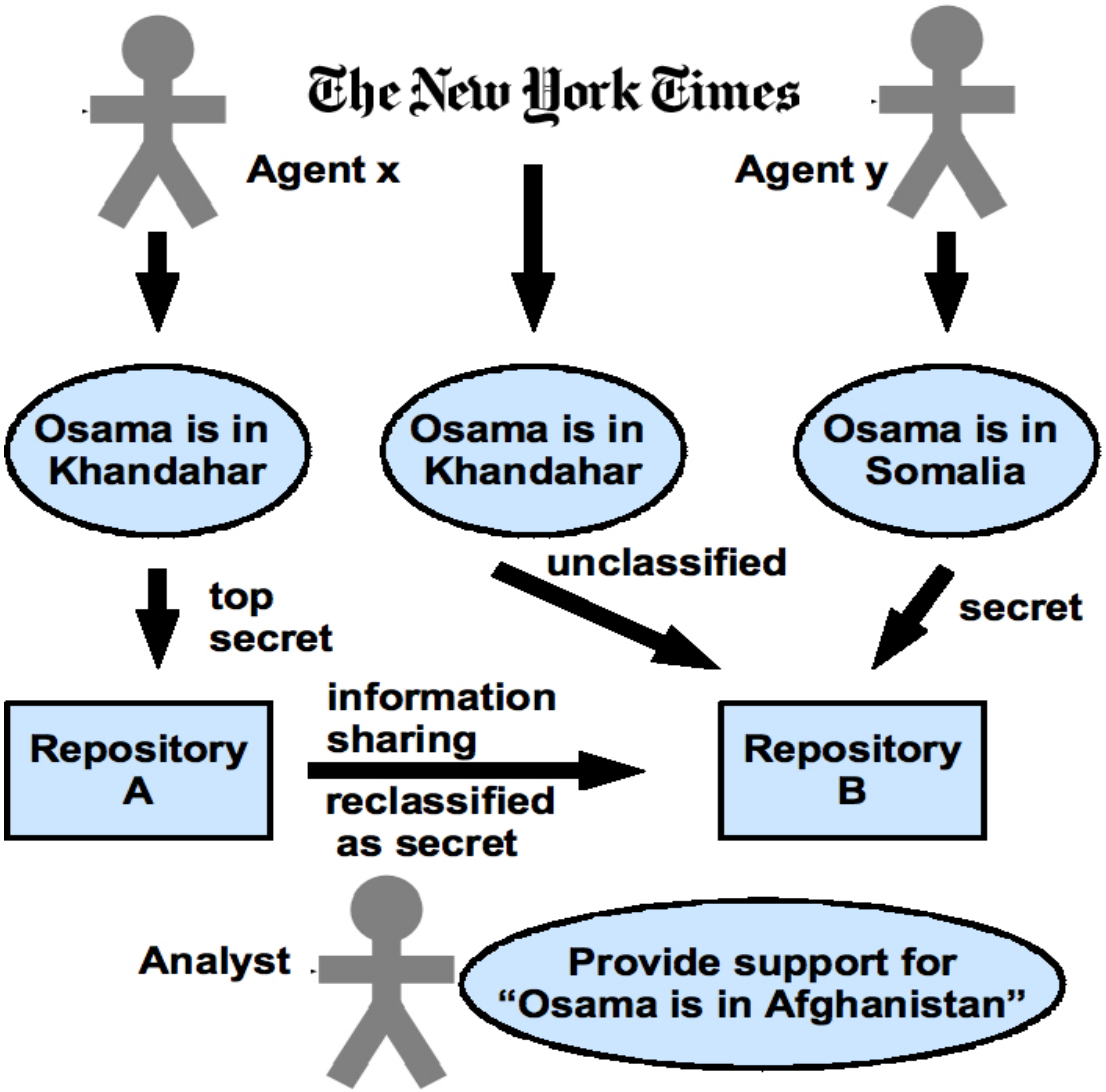
Formalization

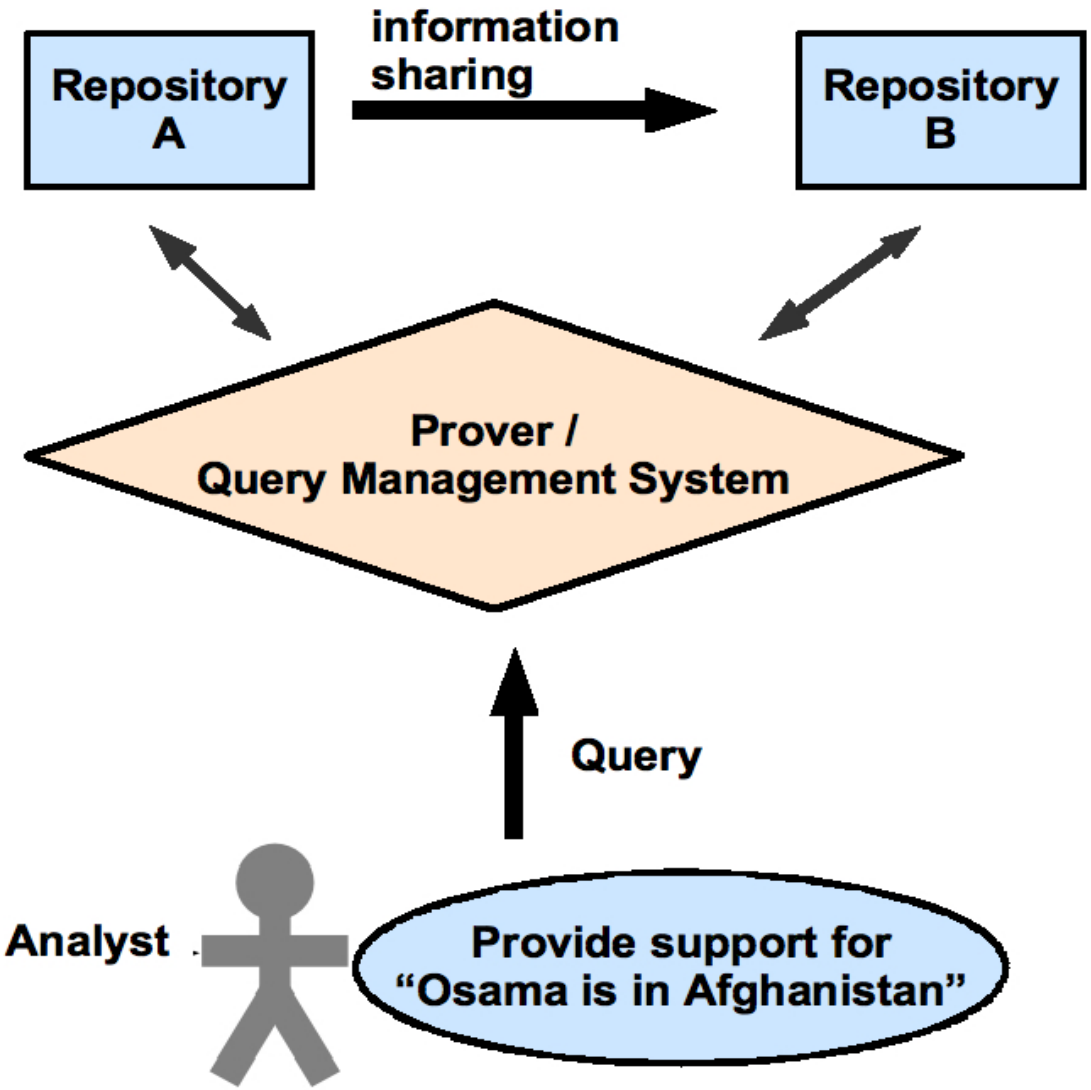
Our Approach

+++DISCLAIMER+++

- Only basic ideas (more details in the paper)
- Theoretical perspective (for implementation ask Bill)

Something was missing




Repository A
information sharing
Repository B
Prover / Query Management System
Query
Analyst
Provide support for "Osama is in Afghanistan"

Tokens and 'meta-data' are treated as first-class entities by the prover

Information about record in repository A:

Record(token001) &
ResidesIn(token001) = repository_A &
PropositionalContent(token001) =
(that (LocatedIn (osama kandahar))) &
ClassifiedAs (token001 top_secret) &
Compartment (token001 al-quaeda_cmpt) &
Source(token001) = agent_x

Provenance info is treated in the same way

Information about record in repository B:

Record(token002) &
ResidesIn(token002) = repository_B &
PropositionalContent(token002) =
(that (LocatedIn (osama kandahar))) &
ClassifiedAs(token002 secret) &
CopyOf(token002, token001) &
ResidesIn(token001) = repository_A

SupportedBy: relation between a proposition and a sequence of tokens

A[s] shorthand for SupportedBy((that A), s)

Axioms:

- Record(x)→SupportedBy(PropositionalContent(x),x)
- $A \rightarrow A[\,]$
- $(A[s_1]\&B[s_2]\&\Diamond(A\&B)) \rightarrow (A\&B)[s_1, s_2]$
- $(A[s]\&\Diamond A\&\Box(A \rightarrow B)) \rightarrow B[s]$

SupportedBy behaves like a paraconsistent strict implication

Examples

- LocatedIn (osama kandahar)[token001]
- LocatedIn(x kandahar)→LocatedIn(x afghanistan)[ ]
- LocatedIn (osama afghanistan)[token001]

ClearedFor: a relation between users and classifications

name(analyst001) = 'Nathan Hale'
ClearedFor(analyst001 top_secret)
ClearedFor(analyst001 asia_cmpt)
ClearedFor(analyst001 terrorism_cmpt)

Analyst's question

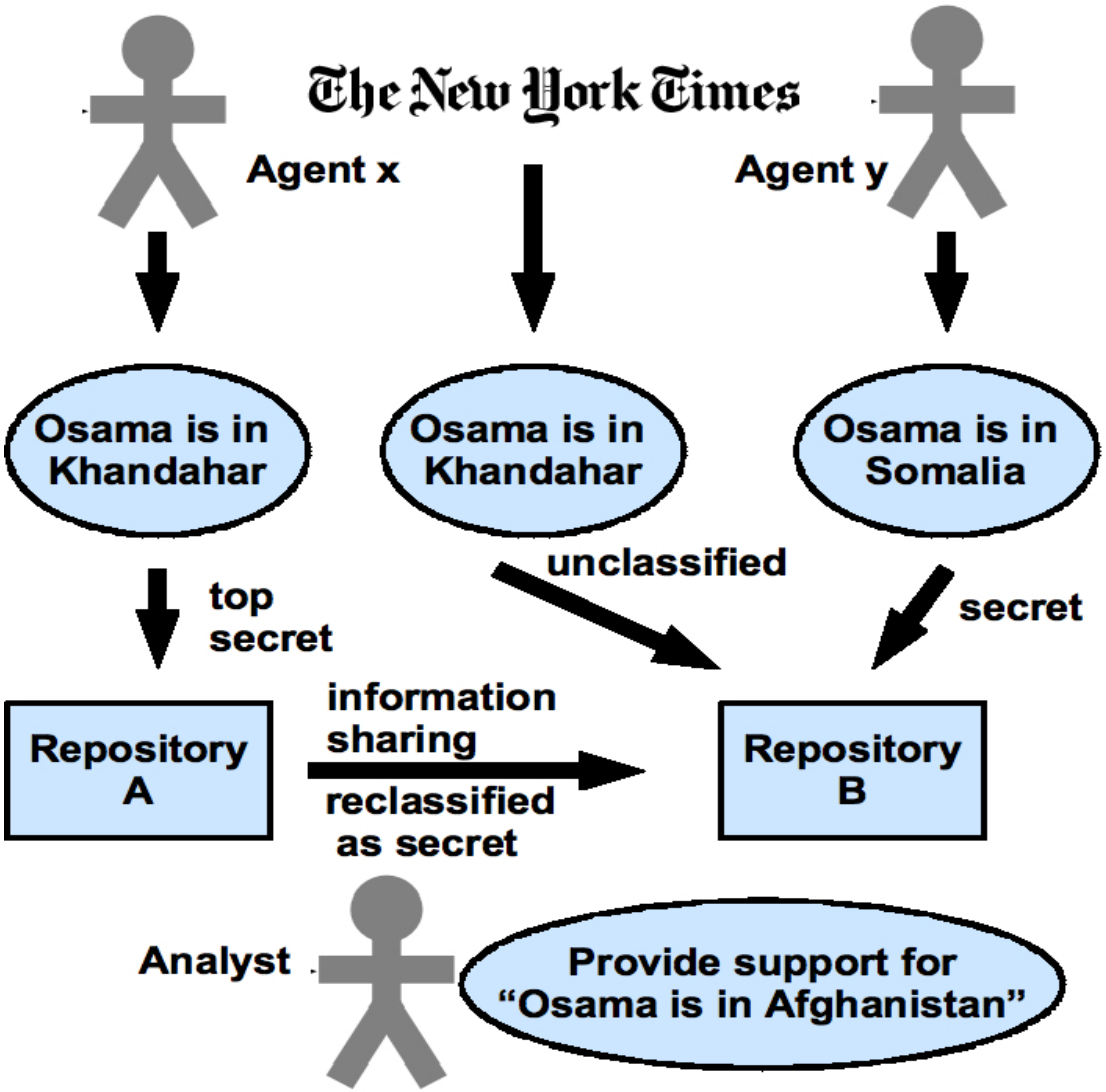
Nathan Hale: Is Osama in Afghanistan?

The system turns it into this query

Find a sequence of tokens s such that

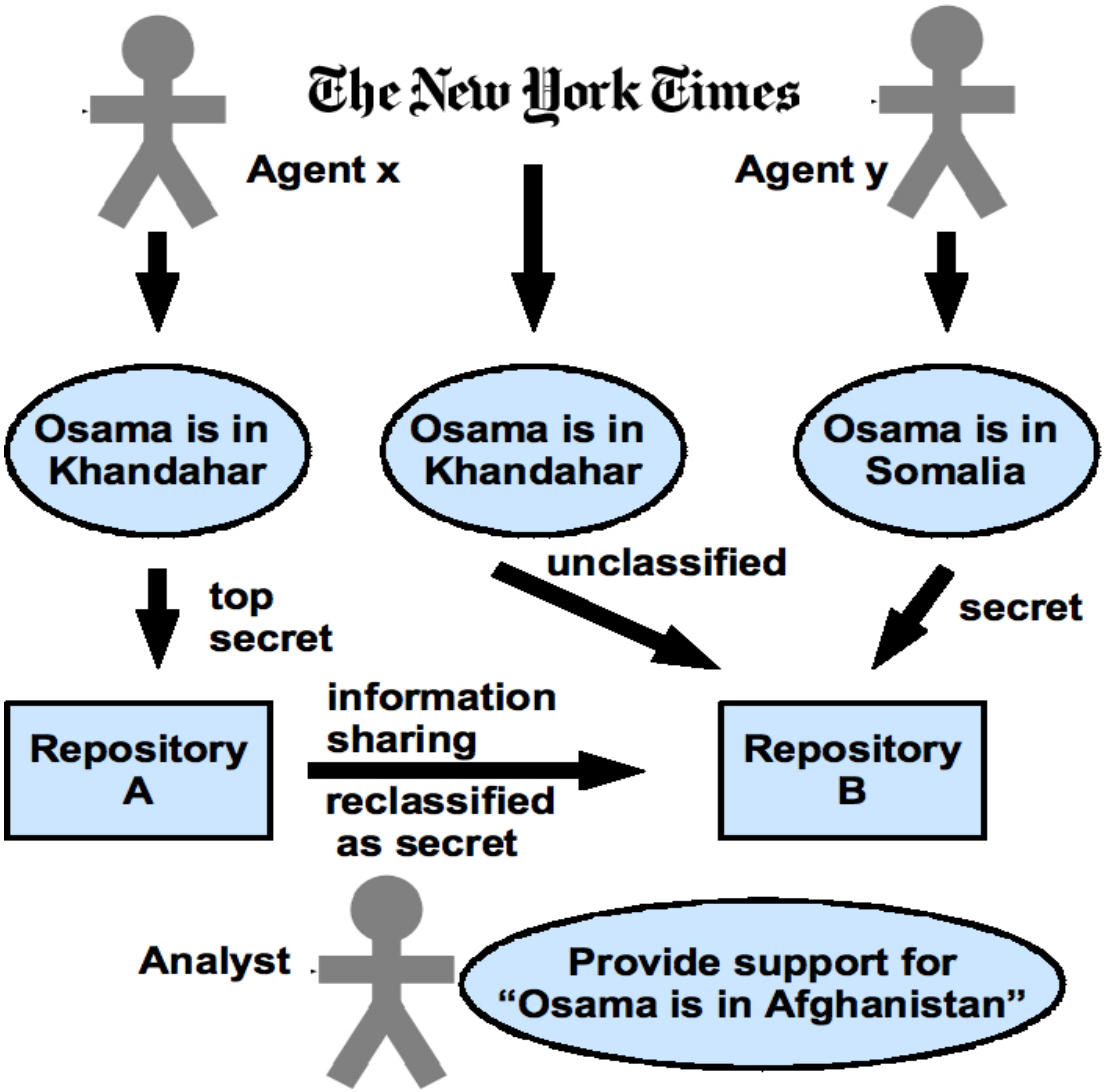
- s supports that Osama is in Afghanistan
- NH is cleared for security levels and compartments of elements of s
- s does not contain an element that is a copy of another token that
  - NH has clearance for, and
  - resides in an accessible repository

Revisiting the Challenges



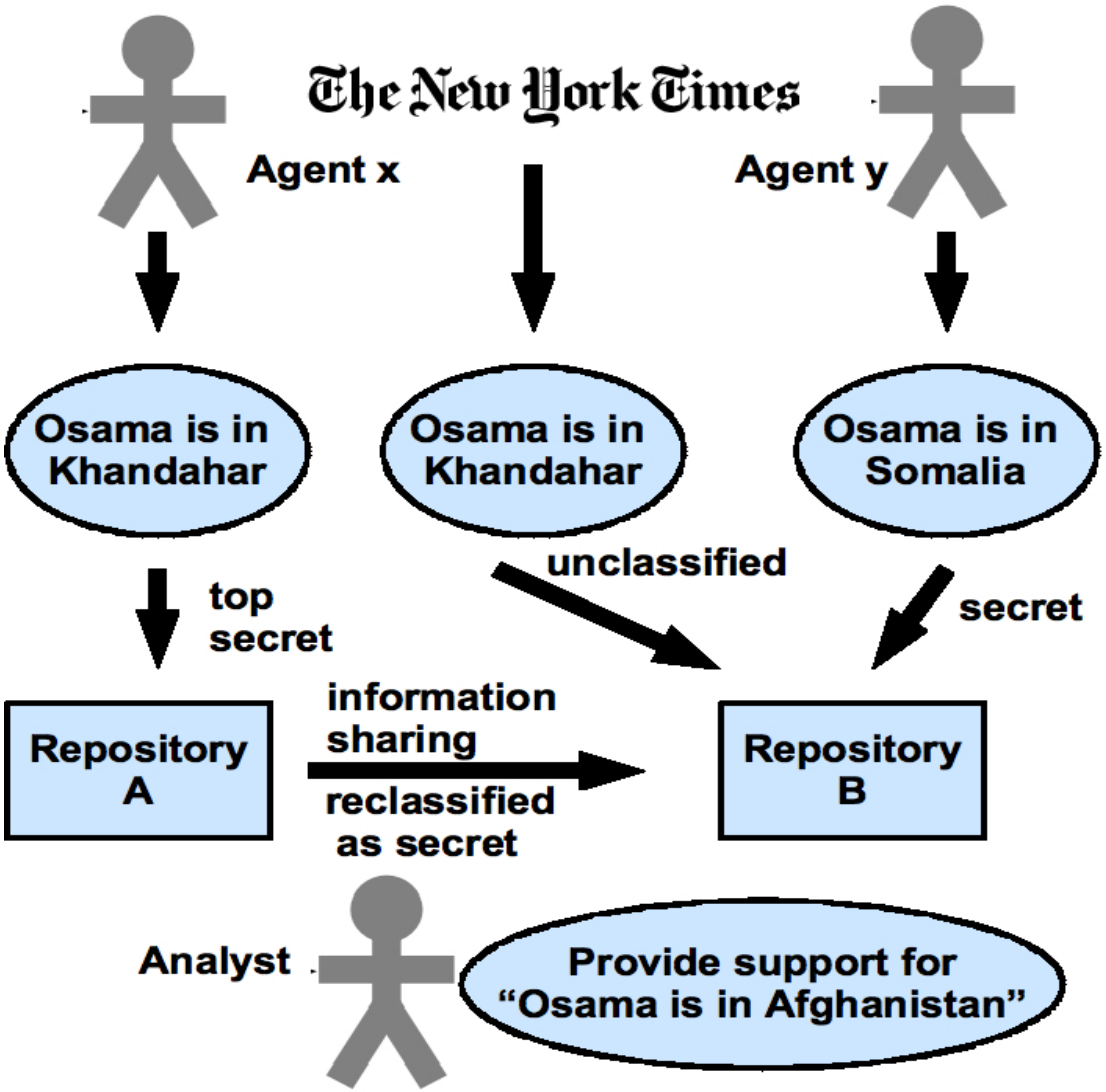
Challenge 1: Reasoning with geospatial information

Revisiting the Challenges



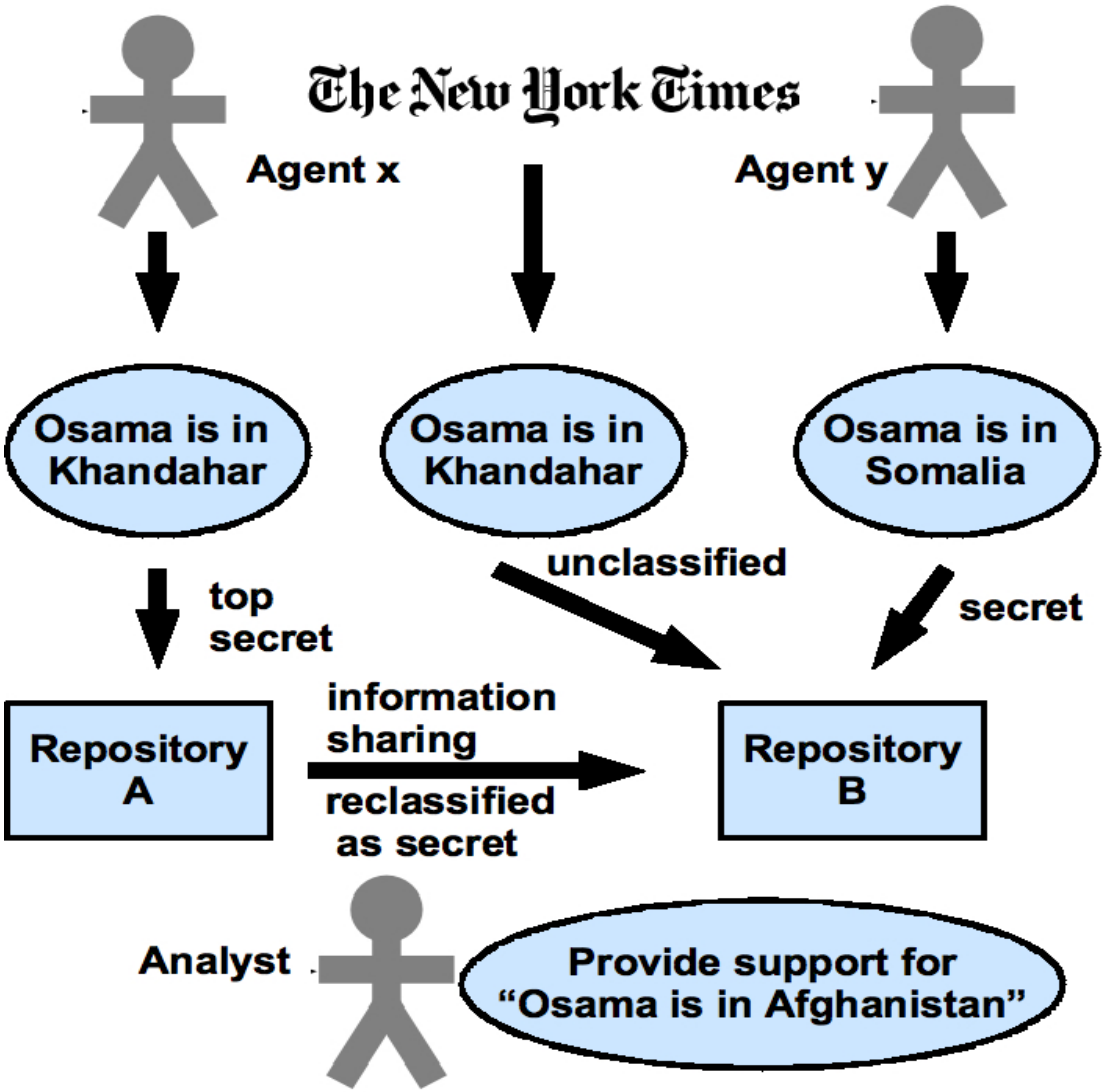
Challenge 2: Reasoning with inconsistent information

Revisiting the Challenges



Challenge 3: Security classification of analyst impacts query

Revisiting the Challenges



Challenge 4: One record in B is a 'copy' of the record in A