

Privacy-Aware and Scalable Content Dissemination in Distributed Social Networks

Pavan Kapanipathi ^{1,2}, Julia Anaya ¹, Amit Sheth ², Brett
Slatkin ³, Alexandre Passant ¹

¹ Digital Enterprise Research Institute, Galway – Ireland

² Kno.e.sis, Dayton, OH- USA

³ Google, San Fransisco, CA - USA

Agenda

- Background
- Motivation
- Semantic Hub – Protocol
- SMOB + Semantic Hub
- Conclusion
- Future Work

Social Networks

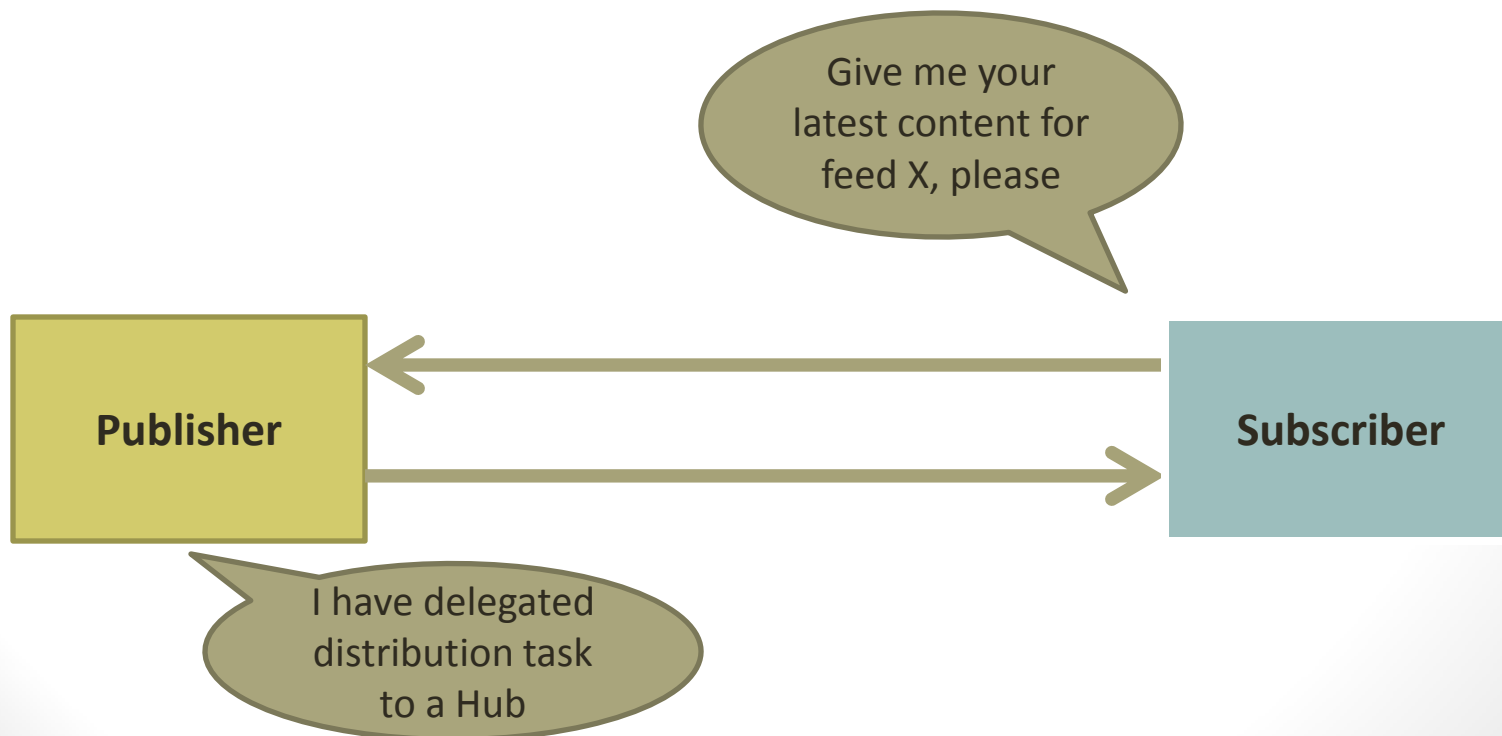
- Centralized Social Networks (CSN)s
 - Twitter and Facebook.
- What's missing?
 - Users don't own their data
 - Consistent privacy and performance concerns
- Distributed Social Networks
 - Diaspora, SMOB, StatusNet, OneSocialWeb

Distributed Social Networks

- Users own their data
 - Own Social Space
 - Decentralized servers
- Better Performance
- Google's PubSubHubbub protocol
 - Communication protocol used in most Distributed Social Networks
 - Improves scalability
 - Near-instant notifications

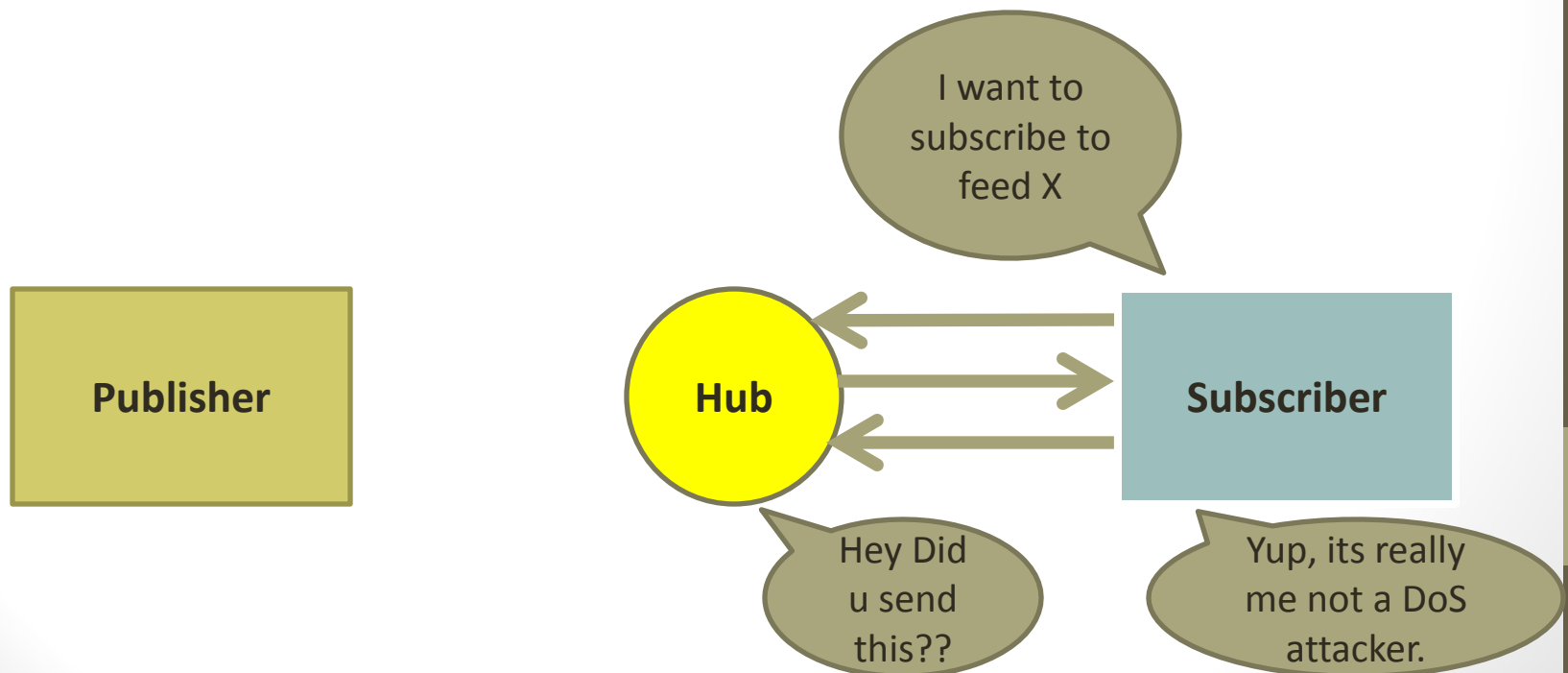
PubSubHubbub

- PubSubHubbub – Communication protocol
 - Simple, Open, web-hook based pubsub protocol
 - Extension to RSS, Atom.



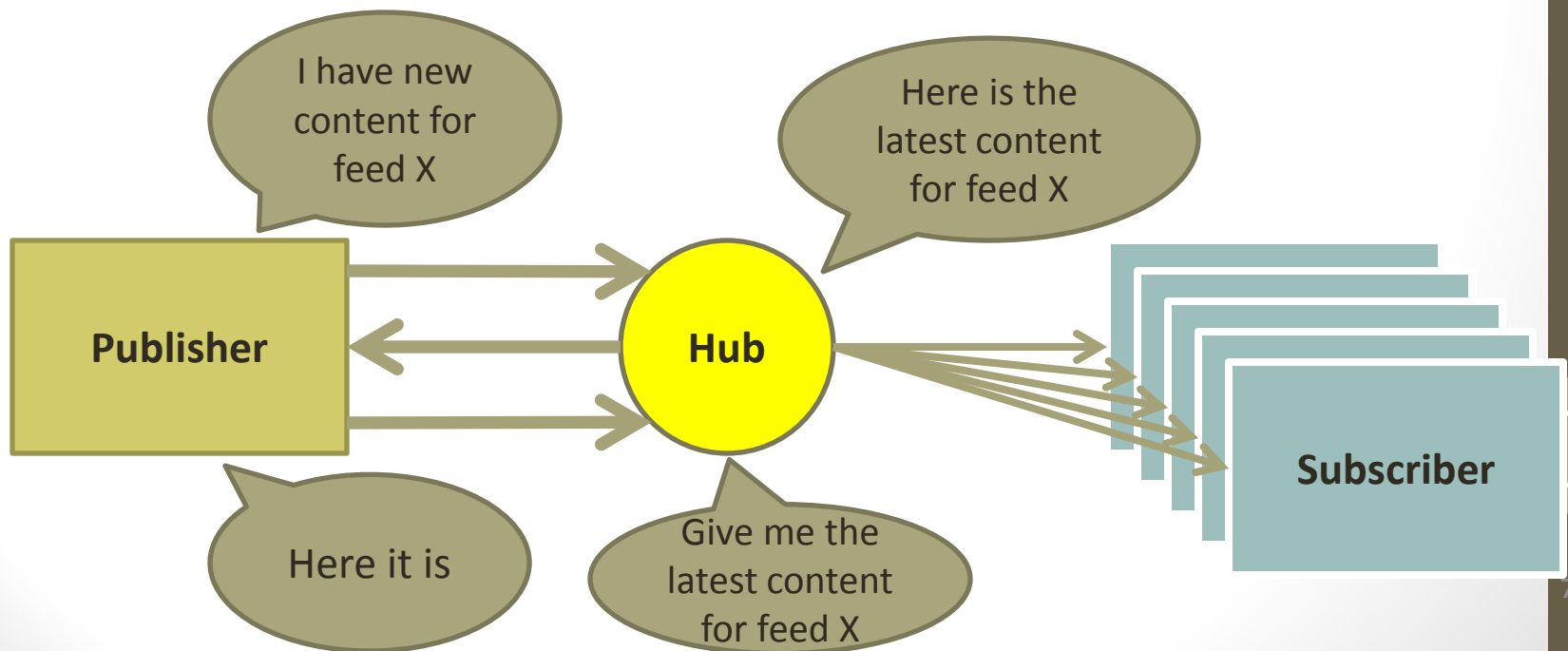
PubSubHubbub

- PubSubHubbub – Communication protocol
 - Simple, Open, web-hook based pubsub protocol
 - Extension to RSS, Atom.



PubSubHubbub

- PubSubHubbub – Communication protocol
 - Simple, Open, web-hook based pubsub protocol
 - Extension to RSS, Atom.



Motivation – Privacy

- Twitter – Data is either
 - Private – authorized followers
 - Public – Everyone (No authorization required to follow)
“Cisco just offered me a job! Now I have to weigh the utility of a fatty paycheck against the daily commute to San Jose and hating the work.”



Why People Unfollow on Twitter

Posted on [June 6, 2011](#) by [Jim Hamill](#)



On Friday 3rd June, I posted the following on my twitter page @DrJimHamill

*What would make you want to 'unfollow' someone on #twitter?
Will post results on our blog.*

Over 50 people replied to the post, either on Twitter itself or on my [Linkedin page](#). A big thank you to everyone who took the time to reply.

While my academic colleagues would argue that 50 is hardly a representative sample of Twitter users, the responses are very interesting nevertheless. The 20 comments listed below are representative of all those received.

Noise, advertising, spam, rudeness, lack of reciprocity, automated tweets, no engagement

I've only ever unfollowed 1 person; because she kept writing self-indulgent nonsense about herself

Also known as Wheaton's Law !! (ps – I had to Google the meaning of this myself 😊 JH)

Negativity, too many entries of irrelevant and wordy content, pushing sales too hard, superficial tweets

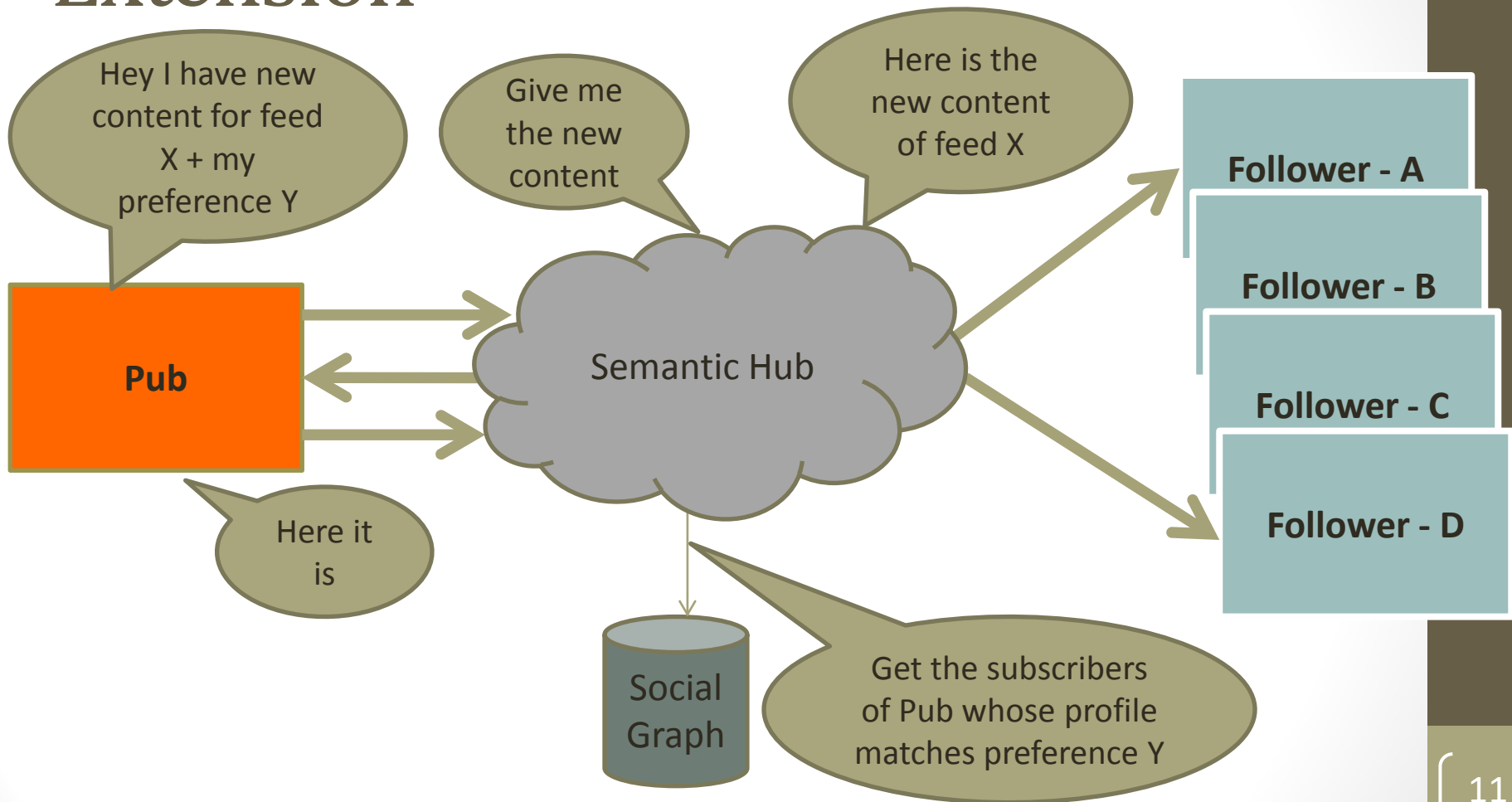
Repeated #FF's – loads of them. Constantly plugging the same site/blog with no new information

Someone with a depressing personality. You'll be surprised how many there are

Protocol – Extension (Overview)

- Publishers' Preferences
 - Content also includes its preferences (recipients)
- Complexity at the Hub
 - Store Social Graph of the Publisher
 - Filter appropriate Subscribers to push the content

PubSubHubbub Protocol Extension



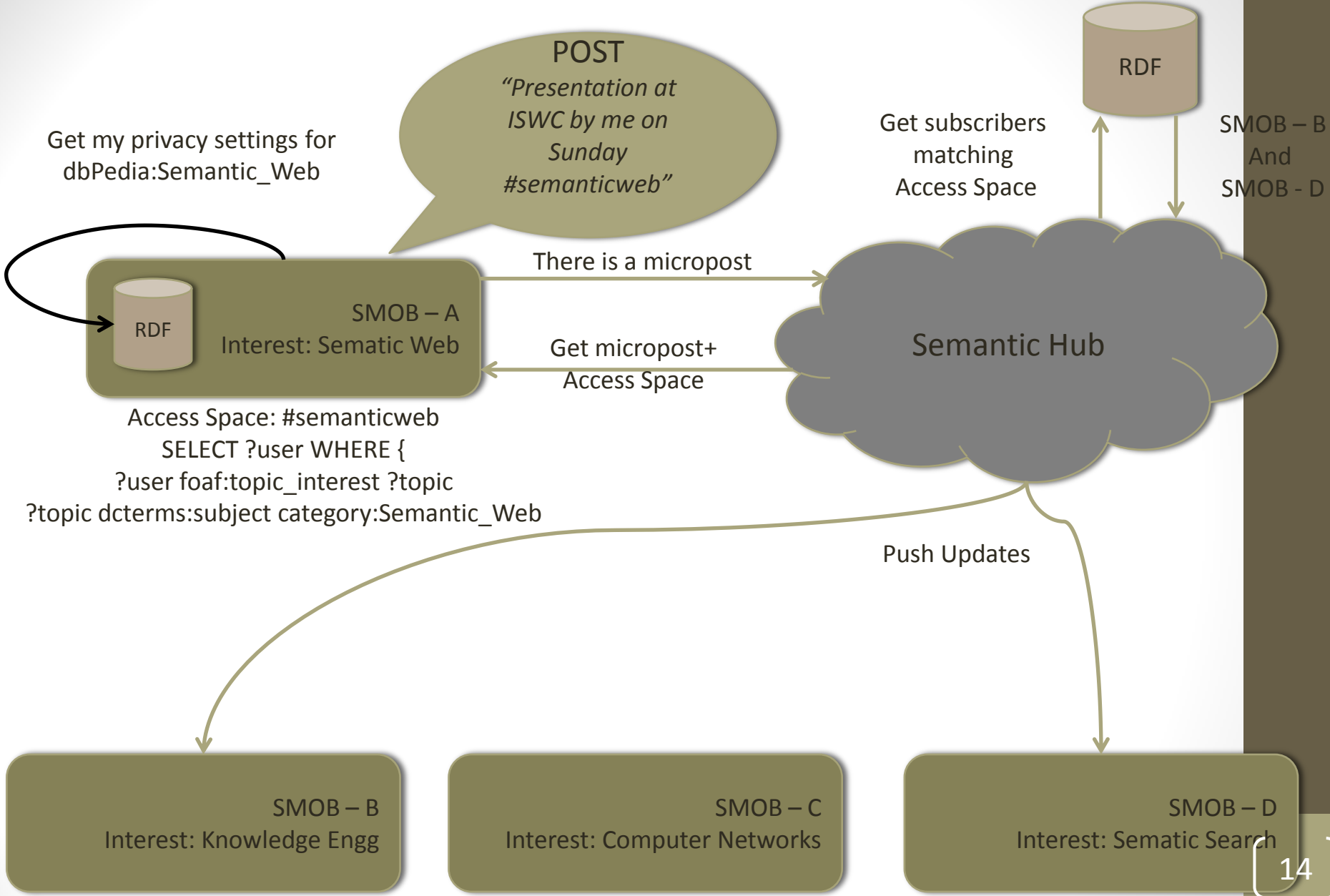
Protocol In-Use

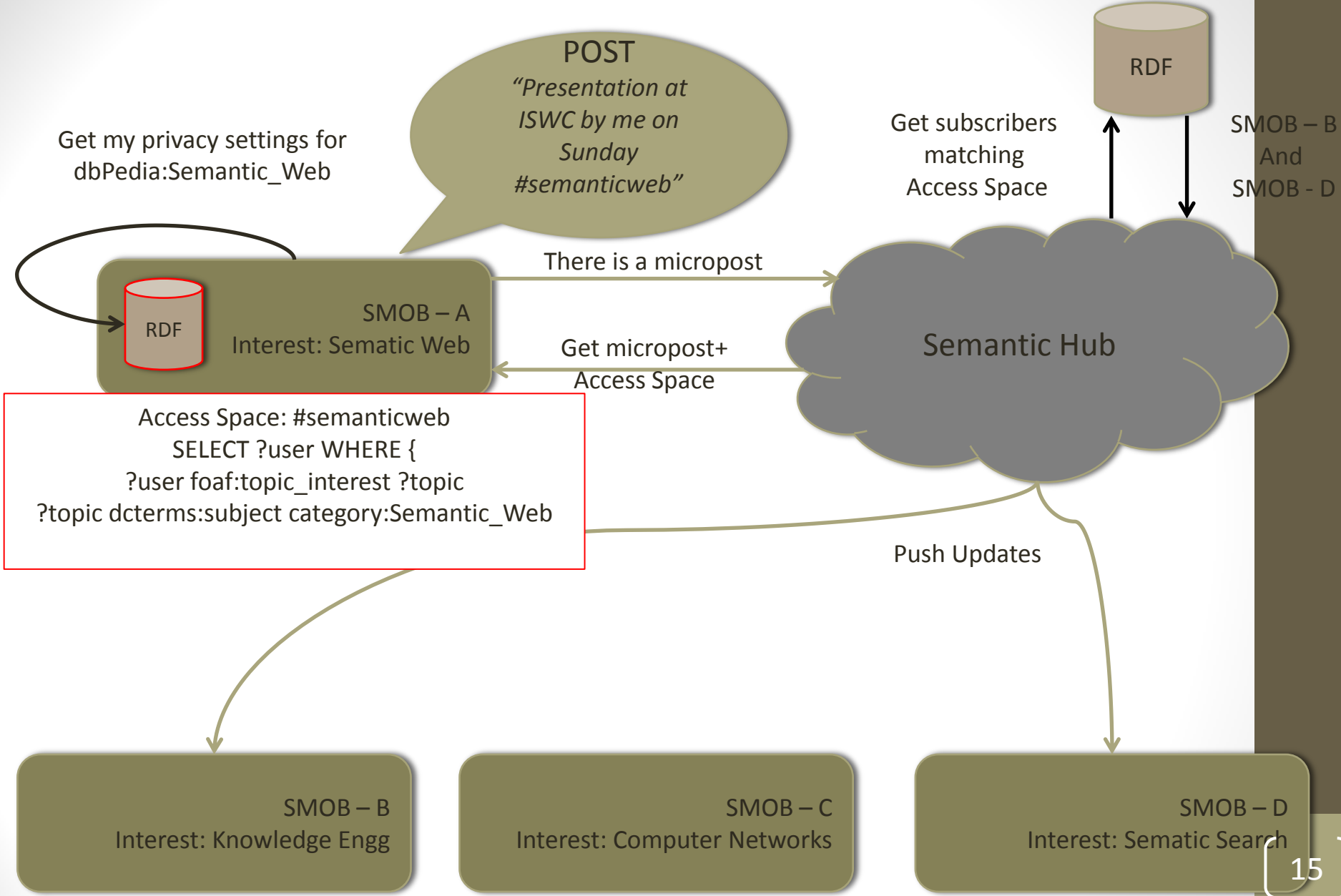
- SMOB
 - Privacy-Aware Microblog dissemination

- Personalized Filtering
 - Twitter Stream (SPIM Workshop)

SMOB + Semantic Hub

- Open and Distributed Semantic Microblogging Framework
 - <http://smob.me>
- PuSH implementation used to broadcast content
 - Privacy was a concern
- Our Privacy-Aware protocol implementation used
 - <http://semantichub.appspot.com>





Publisher – Privacy Preferences

- Feature based preferences at Publisher
 - Content generated for topic Linked Data
 - Micropost containing hashtag #knoesis
- Modeled using Privacy Preference Ontology (PPO)
 - Fine grained restrictions
 - SPARQL Queries representing the subset of subscribers from the Social Graph in the hub

Privacy Preference Ontology (PPO)

- Publishing Data on Web as a part of Linked Open Data has privacy issues.
 - Web Access Control ontology (WAC)
- PPO -- Light weight vocabulary
 - On top of Web Access Control Ontology
 - Fine grained privacy preferences for RDF Data
 - Example: Share email, phone number in my FOAF profile to Users with workplace as Kno.e.sis Center.

Privacy Preferences using PPO

Example: Restrict all the documents tagged with Semantic web to only those users who are interested in Semantic Web

```
<http://example.org/privacy/3> a ppo:PrivacyPreference;
  ppo : appliesToResource
  <http://xmlns.com/foaf /0.1/Document>;
  ppo : hasCondition [
  ppo : hasProperty tag:Tag;
  ppo : resourceAsObject dbpedia : Semantic Web
  ];
  ppo: assignAccess acl :Read;
  ppo : hasAccessSpace [
  ppo:hasAccessQuery "SELECT ?user WHERE {
    ?user foaf : topic interest dbpedia:Semantic Web }"
  ].
```

SMOB's Simple Interface – Privacy Preference Generation

Privacy settings

Hashtag that the microposts must contain

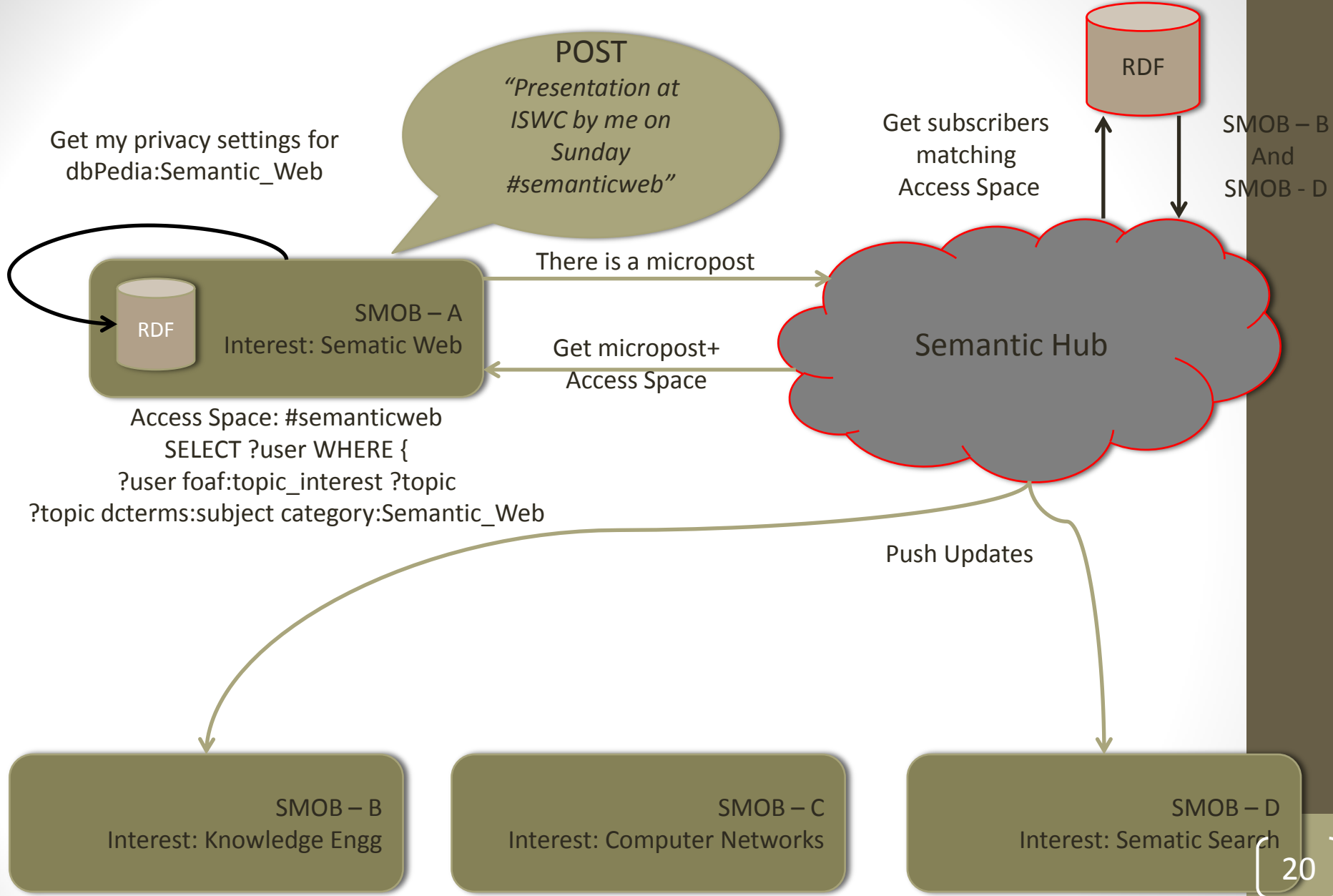
Interest that the subscribers must have to receive the micropost

Relationship that the subscriber must have

Generating Privacy preferences ...

Privacy Preferences generated

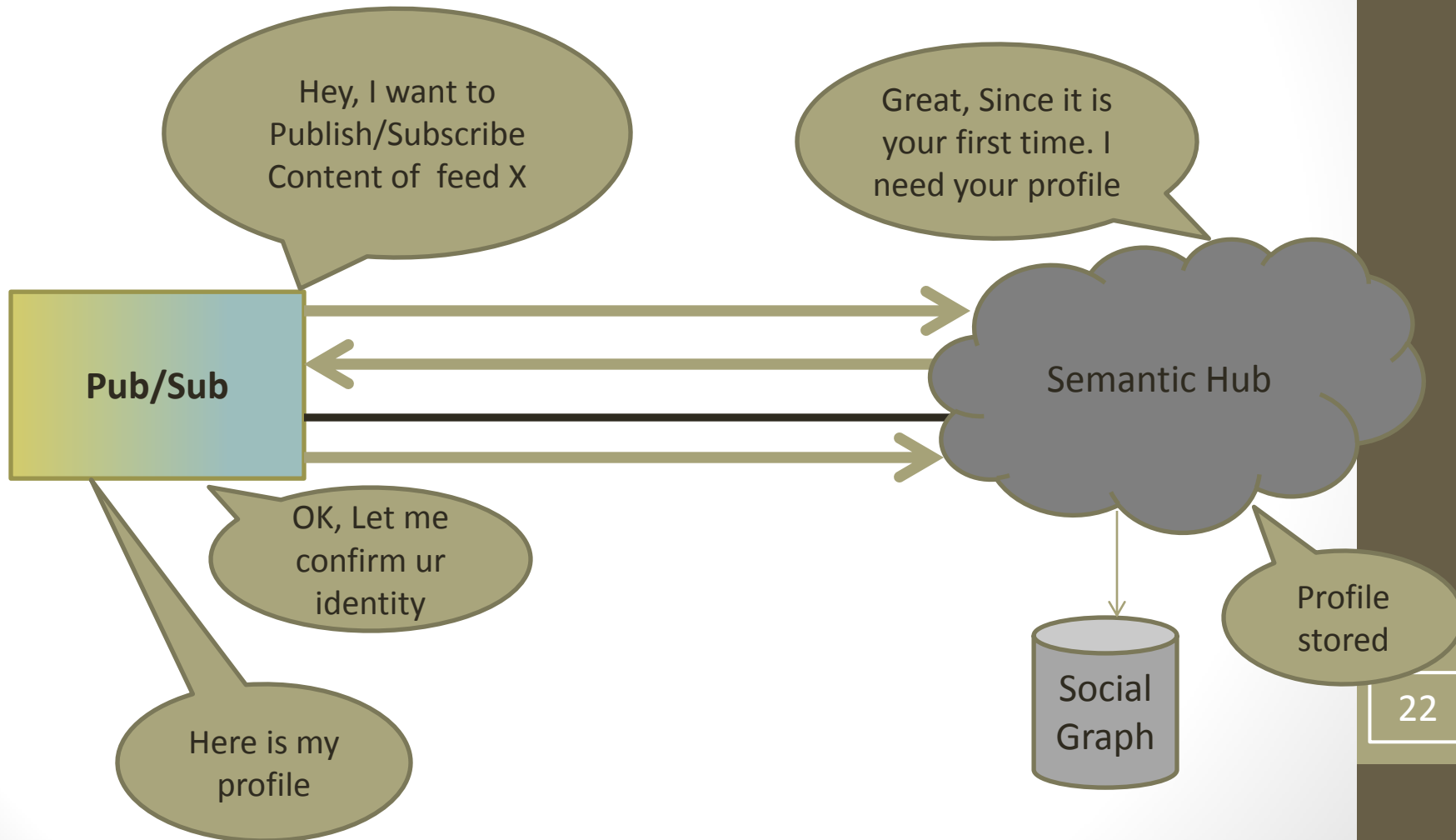
```
http://mysite.org/preference/rdf a ppo:PrivacyPreference;
  ppo:appliesToResource
    http://rdfs.org/sioc/ns#MicroblogPost;
  ppo:hasCondition [
    ppo:hasProperty tag:Tag;
    ppo:resourceAsObject
      http://dbpedia.org/resource/Resource\_Description\_Framework
  ];
  ppo:assignAccess acl:Read;
  ppo:hasAccessSpace [
    ppo:hasAccessQuery "SELECT ?user WHERE {
?user foaf:topic_interest ?topic .
?topic dcterms:subject category:Semantic_Web .}"
  ] .
```



Semantic Hub – Distributed Social Semantic Graph

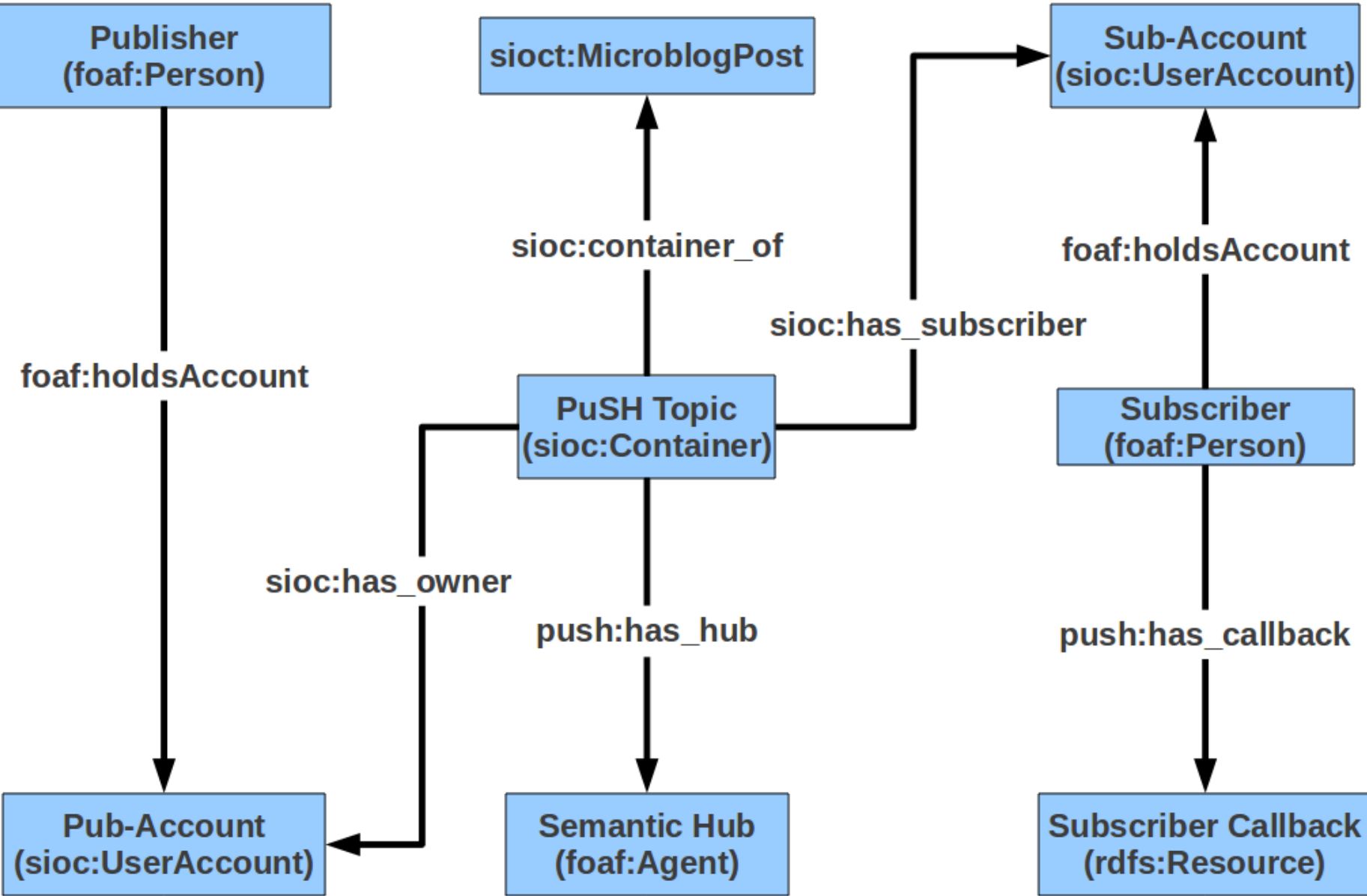
- FOAF Profiles of Publishers/Subscribers
- Semantic Social Graph at the Semantic Hub
 - FOAF profiles linked using push vocabulary
- Distributed
 - Every user owns a Hub

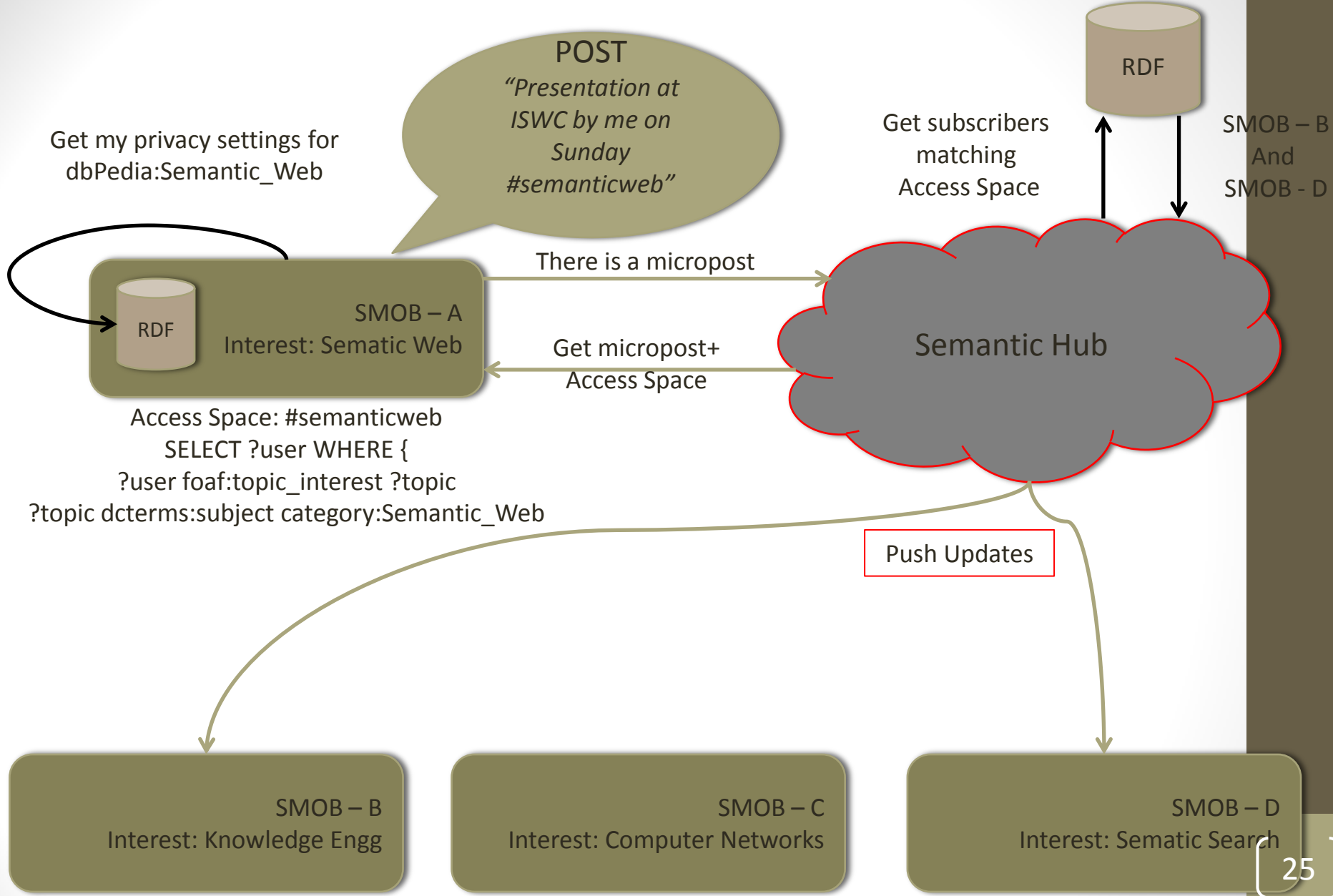
FOAF -- Exchange



Push Vocabulary

- SIOC vocabulary
 - Models information from online communities
 - Communication protocols are not considered
- PubSubHubbub Vocabulary (push)
 - Models information of users communicating via Semantic Hub





Semantic Hub – Content Distribution

- RSS Extension
 - Element to include the privacy preference
 - privacy, accessspace
- Each updated content is received by restricted subscribers
 - Privacy Preference for each content fetches the matched subscribers
 - Preferences are deleted before distribution

RSS Extension

```
<item>
  <title>Only Friends</title>
  <description>
    Send this to only people I know and interested in Semantic
Web
  </description>
  <link>http://example.org /rss</link>
  <guid>123123123123</guid>
  <pubDate>March 06 2001</pubDate>
  <privacy>
    <accessspace>
      SELECT ?user WHERE {
        foaf :me foaf :knows ?user .
        ?user foaf : topic interest dbpedia:Semantic Web
      .
    }
    </accessspace>
    ...
  </privacy>
</item>
```

Personalized Filtering of the Twitter Stream

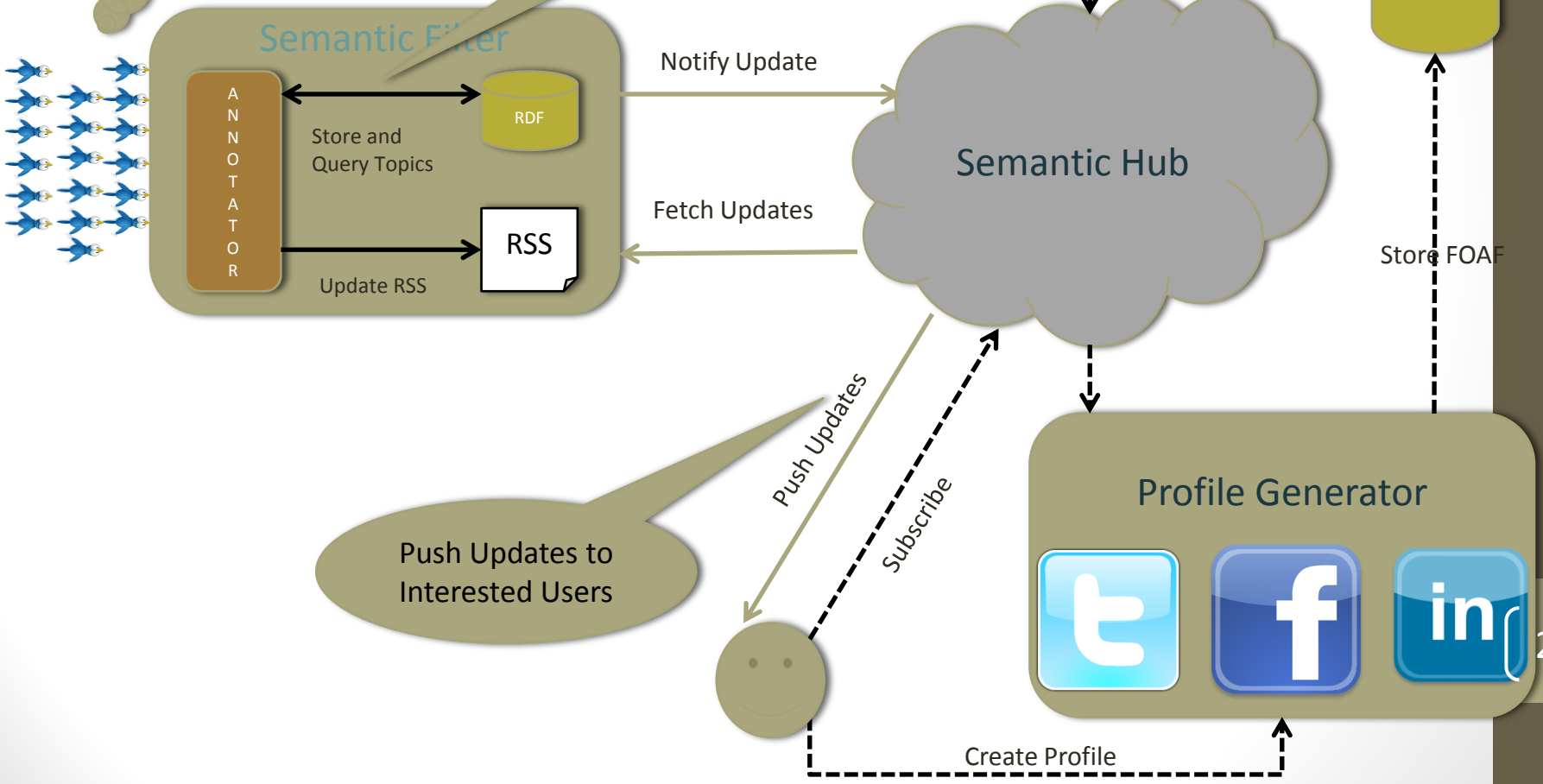
- Profile Generator
 - Automatic generation of User Profiles
- Semantic Filter
 - Annotating Twitter Stream with concepts from Linked Open Data
- Semantic Hub
 - Delivering tweets to appropriate Interested Users (near real-time)

Structure

The new **iPhone** has a 3.5-inch screen, released today

Annotate: iPhone
?user foaf:interest dbPedia:iPhone
Union
?user foaf:interest Category:Apple

Get Subscribers based on preference



Push Updates to Interested Users

Conclusion

- Extension of Google's PuSH protocol to cope with privacy concerns.
- Implementation of the protocol
- Use in SMOB (Microblogging)
- Push vocabulary to model activities via PuSH protocol

Future Work

- Enabling the architecture for mobile devices.
- Communicate to and from devices that are offline and still need to be notified.

Thanks

Acknowledgements:

Fabrizio Orlandi

Owen Sacco

Contact us:

{pavan, amit}@knoesis.org

{alexandre.passant, julia.anaya}@deri.org

brettslatkin@google.com

PubSubHubbub

- PuSH is a scalable protocol, and Google provides a public hub that people can use to broadcast their content.
- This public hub delivers for approximately 40 million unique active feeds
- 117 million subscriptions.
- In two years, approximately 5.5 billion unique feeds have been delivered
- fetching 200 to 400 feeds
- delivering 400 to 600 of them per second.
- Its largest subscribers get between 20 and 120 updates per second from the hub.

