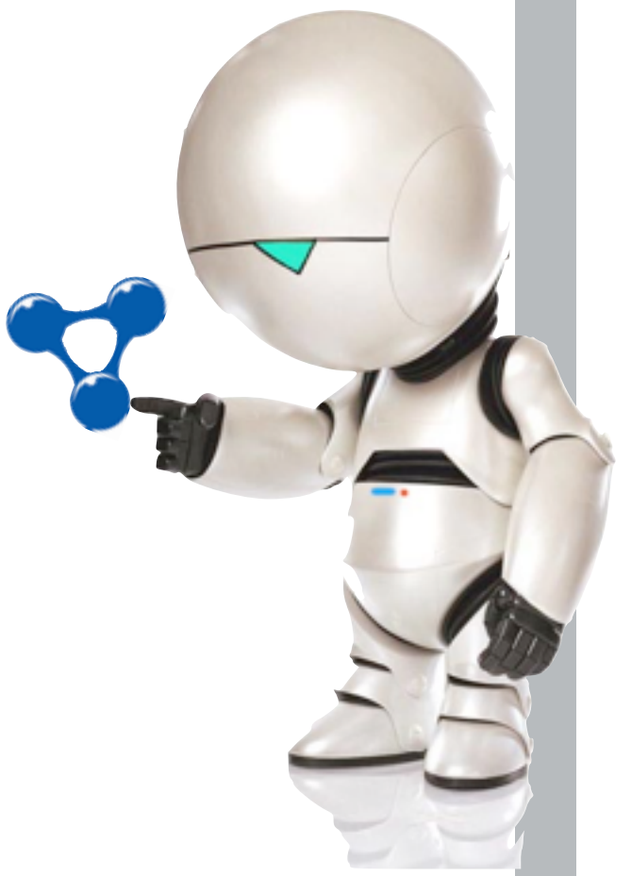
HPI Hasso Plattner Institut

IT Systems Engineering | Universität Potsdam



Collaboratively Patching Linked Data

A Patch Repository for Linked Datasets

Magnus Knuth, Johannes Hercher,
and Harald Sack
Hasso Plattner Institute, University of Potsdam

USEWOD Workshop @ WWW 2012
April 17, 2012 - Lyon, France

Outline

- Introduction
- Patch Request Ontology
- Architecture / Workflow
- Use Case
  - WhoKnows?
  - Patch Repository
- Outlook

Problem

- the Web of Data is noisy
- dataset not under employers control
- erroneous Data distributed to multiple local data stores
- missing error correction propagation mechanisms
  - for commits, for updates

no read-write-web

performance
stability
integration

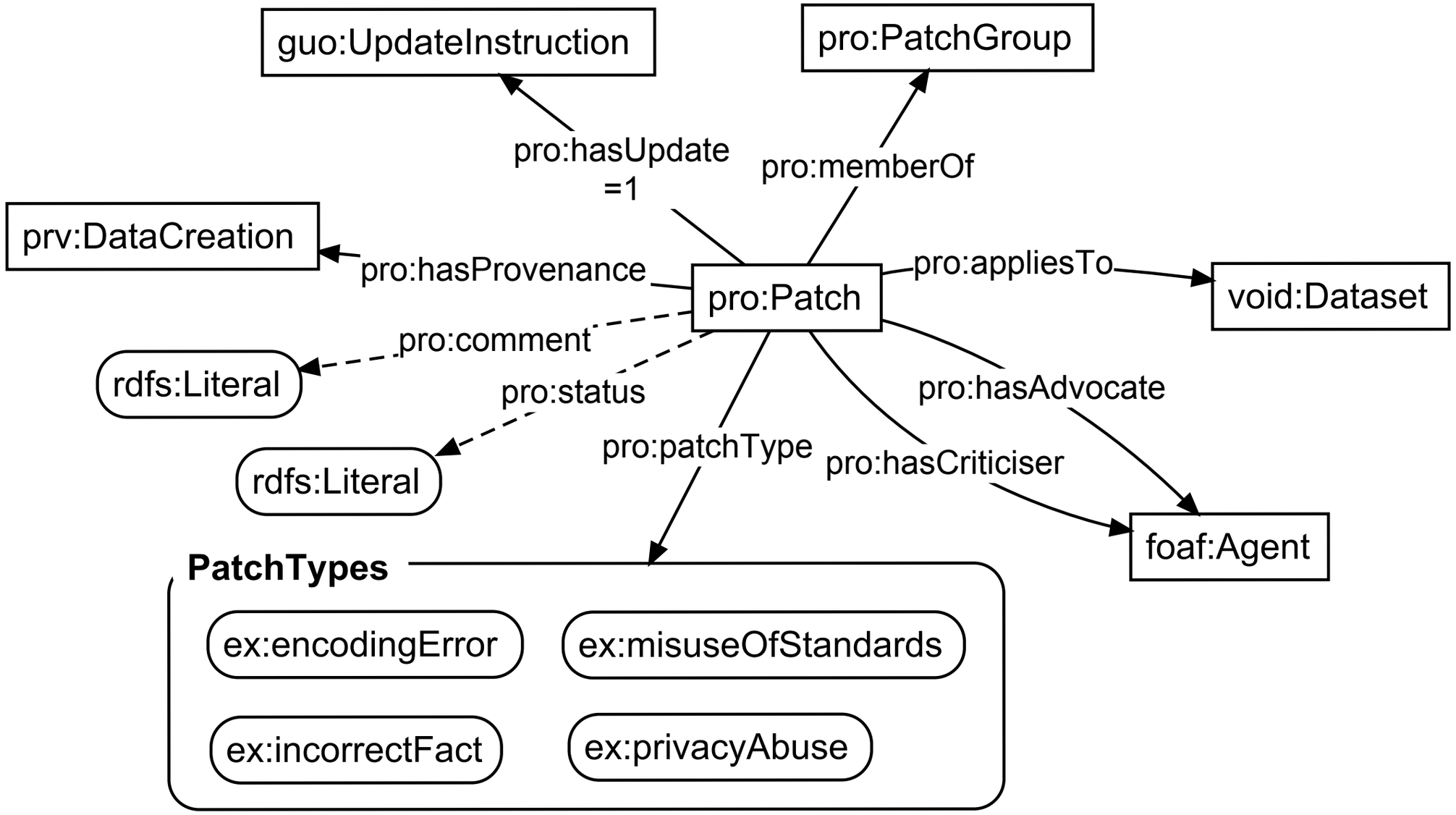
- examples from DBpedia:
  - ```
    dbp:Ukraine dbo:anthem dbp:Transliteration,
                           dbp:Ukrainian_language .
    ```
  - ```
    dbp:Fred_Records dbo:distributingCompany dbp:Japan,
                                             dbp:United_States,
                                             dbp:United_Kingdom .
    ```
  - ```
    dbp:Rhode_Island dbo:language dbp:De_jure,
                                  dbp:De_facto .
    ```
    ```

Goals

- a solution that addresses Linked Data error corrections
- an ontology to describe error corrections for Linked Datasets
- a framework to collect Linked Data Patches in a collaborative way
  - explicitly involving data consumers
- a process to propagate Linked Data Patches over multiple local stores

Patch Request Ontology

- *„a simple vocabulary to share data about erroneous triples“*
  - `http://purl.org/hpi/patchr`

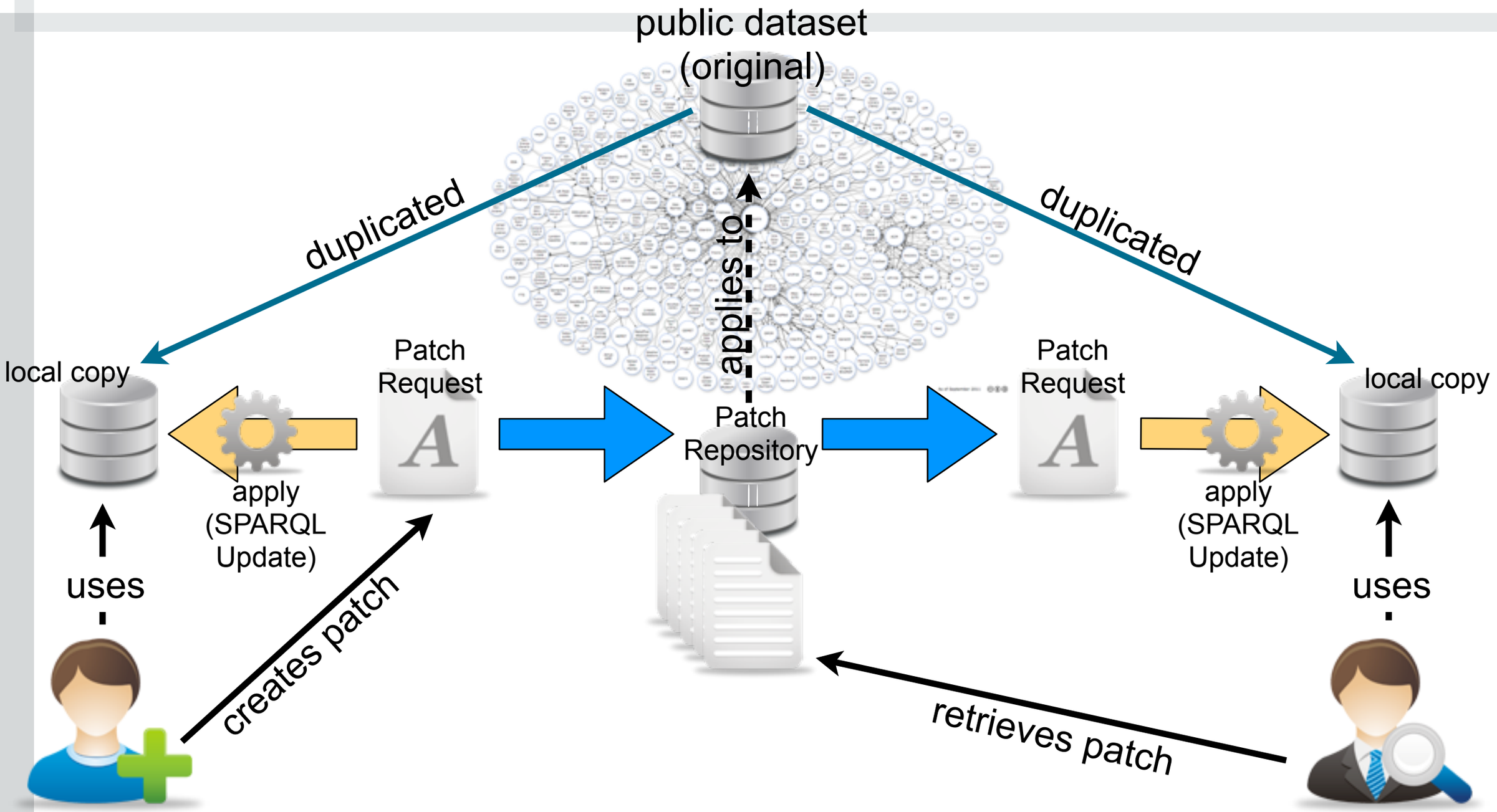


A Patch Request

Patch for exactly one triple

```
repo:Patch_15  a  pro:Patch  ;
  pro:hasUpdate  [
    a  guo:UpdateInstruction  ;
    guo:target_graph  <http://dbpedia.org/>  ;
    guo:target_subject  dbp:Oregon  ;
    guo:insert  [
      dbo:language  dbp:English_language  ]
    ]  ;
  pro:hasAdvocate  repo:Player_25  ;
  pro:appliesTo  <http://dbpedia.org/void.ttl#DBpedia>  ;
  pro:status  "active"  ;
  pro:hasProvenance  [
    a  prv:DataCreation  ;
    prv:performedBy  repo:WhoKnows  ;
    prv:involvedActor  repo:Player_25  ;
    prv:performedAt  "..."^^xsd:dateTime  ]  .
```

Architecture / Workflow



Workflow

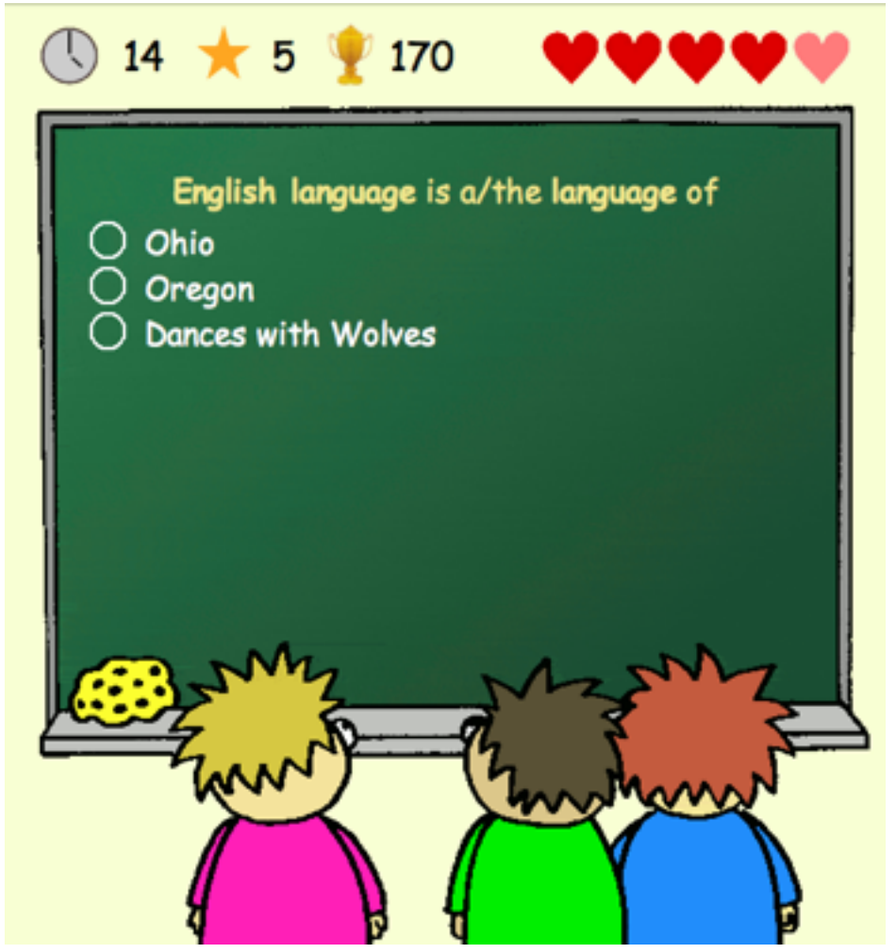
1. find an error
   - by human user or automatic algorithm
2. create a patch
3. commit to (central) repository
   - there should be one responsible repository for each dataset
   - if patch preexists: one more vote
4. other users / dataset providers retrieve patches from repository
   - via SPARQL query
   - customizable to individual requirements
5. apply updates to local dataset
   - easy transformation of patch request to SPARQL Update query

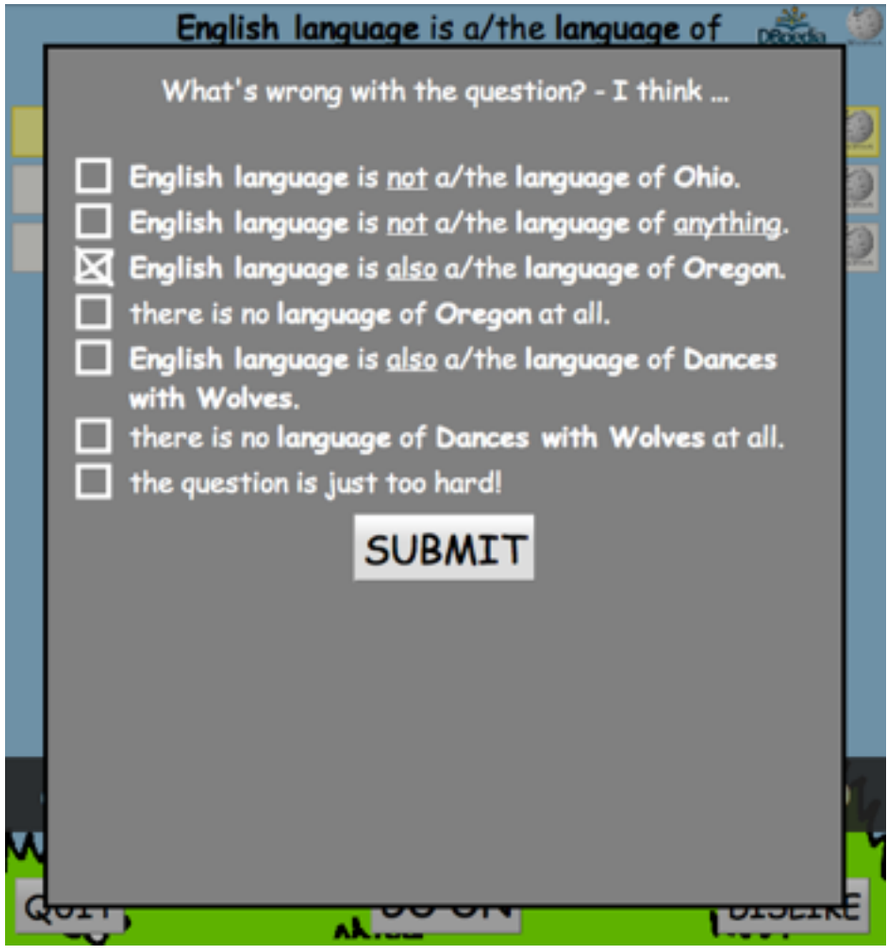
Use Case: *WhoKnows?*

- GWAP generates multiple choice questions from DBpedia facts
- player identifies wrong triples if the question (or desired answer) makes no sense
- generating patch from user vote

DEMO

http://141.89.225.43/game.html





Patch Repository

- list most recent / most popular patches, individual filtering
- show patches for individual resources

DEMO

http://141.89.225.43/patchr/browse.php

Filtered Patch Requests

action: -- Please select --

minimum Advocates: 3

limit: 500

maximum Criticisers:

Submit

Result size: 7.

action	advocates	criticisers	subject	predicate	object	graph	last time
insert	6	1	dbp:Alaska Links: DBpedia, Wikipedia	dbo:language	dbp:English_language	http://dbpedia.org/	2012-03-07 16:40:31
delete	5	1	dbp:Mexico Links: DBpedia, Wikipedia	dbo:regionalLanguage	dbp:Languages_of_Mexico	http://dbpedia.org/	2012-03-07 16:37:03
insert	5	1	dbp:Nevada Links: DBpedia, Wikipedia	dbo:language	dbp:English_language	http://dbpedia.org/	2012-02-27 23:43:39
insert	5	1	dbp:Oregon Links: DBpedia, Wikipedia	dbo:language	dbp:English_language	http://dbpedia.org/	2012-03-07 16:40:31
insert	3	1	dbp:Texas Links: DBpedia, Wikipedia	dbo:language	dbp:English_language	http://dbpedia.org/	2012-02-27 23:42:06
insert	3	1	dbp:102nd_Intelligence_Wing Links: DBpedia, Wikipedia	dbo:country	dbp:United_States	http://dbpedia.org/	2012-02-27 09:39:41
insert	3	1	dbp:Arizona Links: DBpedia, Wikipedia	dbo:language	dbp:English_language	http://dbpedia.org/	2011-12-03 18:04:35

© 2011-2012 Hasso Plattner Institute. All rights Reserved.
Website template by Arcsin
Join Us! | Imprint | Page Top ↑

LOD Benefits

- collecting patches from crowdsourcing or algorithmic data curation systems
- providing patches for replicated Linked Datasets
  - improving data quality
  - measuring data quality
- sustainability (Use Case: DBpedia): fix errors at their source, i.e. Wikipedia

Outlook

- effective synchronization of patches
- further standardization
- dataset quality evaluation
- API to submit patches
  - validity checking
- advanced trust and access control mechanisms
  - rating patches (vote up/down)
  - provide feedback (comments)
  - reputation management
- pingback to inform dataset providers

Thanks for your attention!

http://purl.org/hpi/patchr-repository