

Utilizing Open Data for interactive knowledge transfer



qKAI (qualifying Knowledge Acquisition and Inquiry)
PhD research project at the Leibniz University of Hanover

Faculty of Electrical Engineering and Computer Science
Institute of Systems Engineering
System- and Computer Architecture

Outline

- Introduction
 - What is the Social Semantic Web (S2W)?
 - How do we define the term Open Data?
 - Which role plays user interaction in here?

- Motivation and claims
 - How to use Open Data in higher-layered applications?

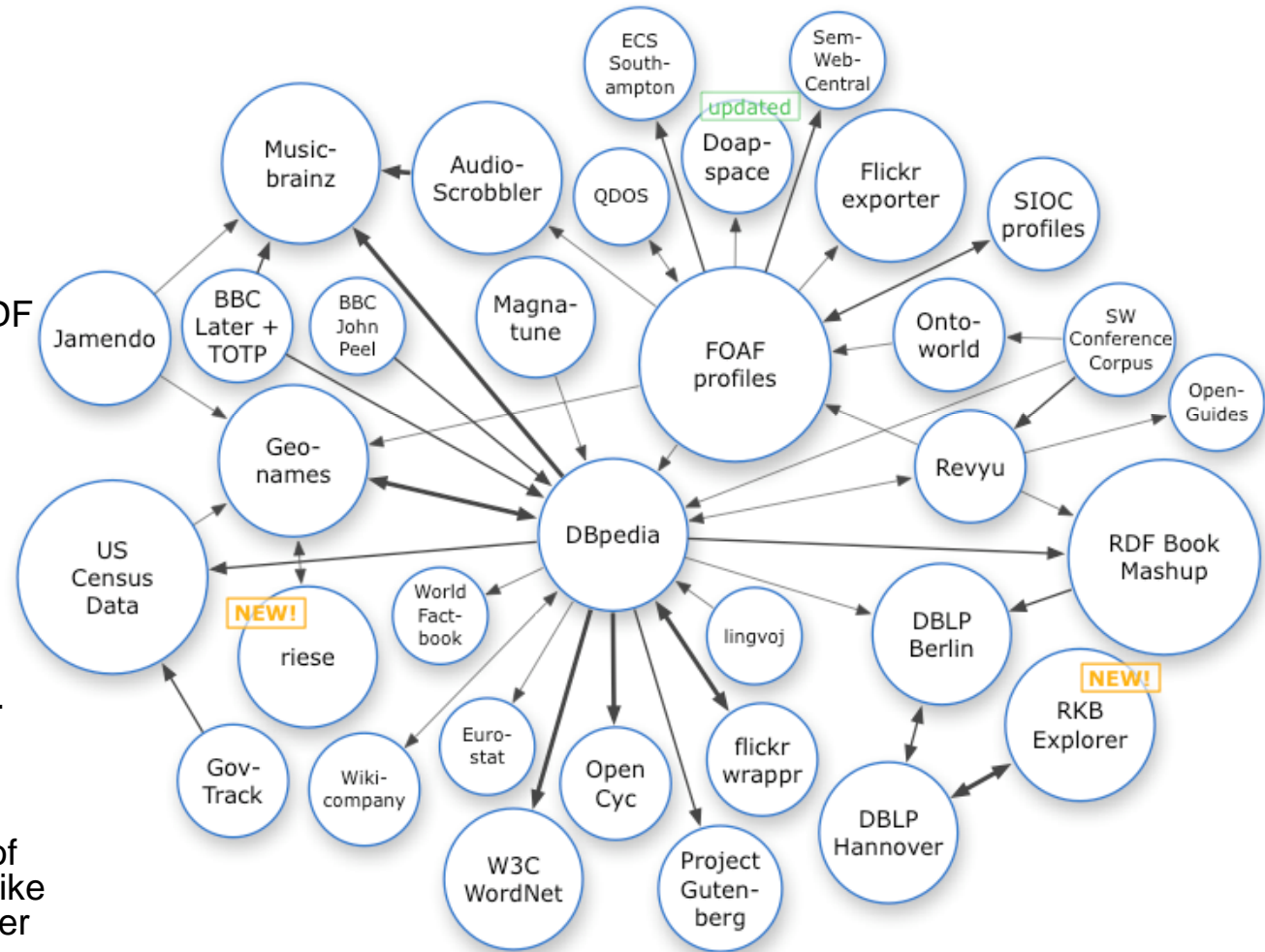
- Refining the concept:
 - Main requirements: resource handling, lightweight user interaction, quality of content.

- Main contribution and focus:
 - The qKAI application framework serves as conceptual basis and system specification.
 - A hybrid data layer embeds distributed resources and non redundant enrichment.
 - Incentive interaction with Open Data is intended on the example of knowledge games: instancing gaming components and sequences.

- Summary and outlook

Introduction: Linked Open Data (LOD)

- "A piece of knowledge is open if you are free to use, reuse, and redistribute it." (Open Knowledge Foundation)
- Knowledge bases are numerous available in RDF as huge LOD in SW, partially linked with each other,
- querying through SPARQL (**S**imple **P**rotocol and **R**DF **Q**uery **L**anguage, W3C standard since 2008).
- Often redundant content in two representations as e.g. (X)HTML and RDF
 - Wikipedia, Freebase, Dbpedia.
- Additionally there are lots of resources outside of LOD like (X)HTML, databases, further wikis, multimedia, ...



Introduction: the role of interaction

- Without interaction Open Data remains futile for the user. Obviously we need some kind of fancy interaction to deploy Open Data for further purpose.
- Examples for interaction with Open Data:
 - Semantic browsing, searching, grouping:
 - Twine, Freebase, Dbpedia, Sindice, Powerset, Swoogle, Stumpedia.
 - Rating and ranking:
 - Revyu.
 - SPARQL query tools
 - very technical, little intuitive.



Freebase: a whole new way of looking at data

Architecture
9 bases made from:

Featured Architecture bases

Washington DC
840 topics including 840 in Museums in Washington, DC

Make a base
Make a Base about Architecture or anything else!

Name your base:

More Bases

- Neutra
- ViceBase
- California History
- Vancouver
- Roaring Twenties
- more ▾

How is Freebase different from the traditional way of looking at data?
Everything in traditional way of looking at data is in a single view. In Freebase, you can build user-defined views to your data. You can pattern the things you see.

What is a base?
A base is a new way of looking at data. You can build user-defined views to your data. You can pattern the things you see. It's a collection of subject. Learn more.

➔ **Searching, displaying, editing, annotating and grouping of content.**

➔ **There is still little continuative advantage without much Incentive for the user to interact and deduce new knowledge.**

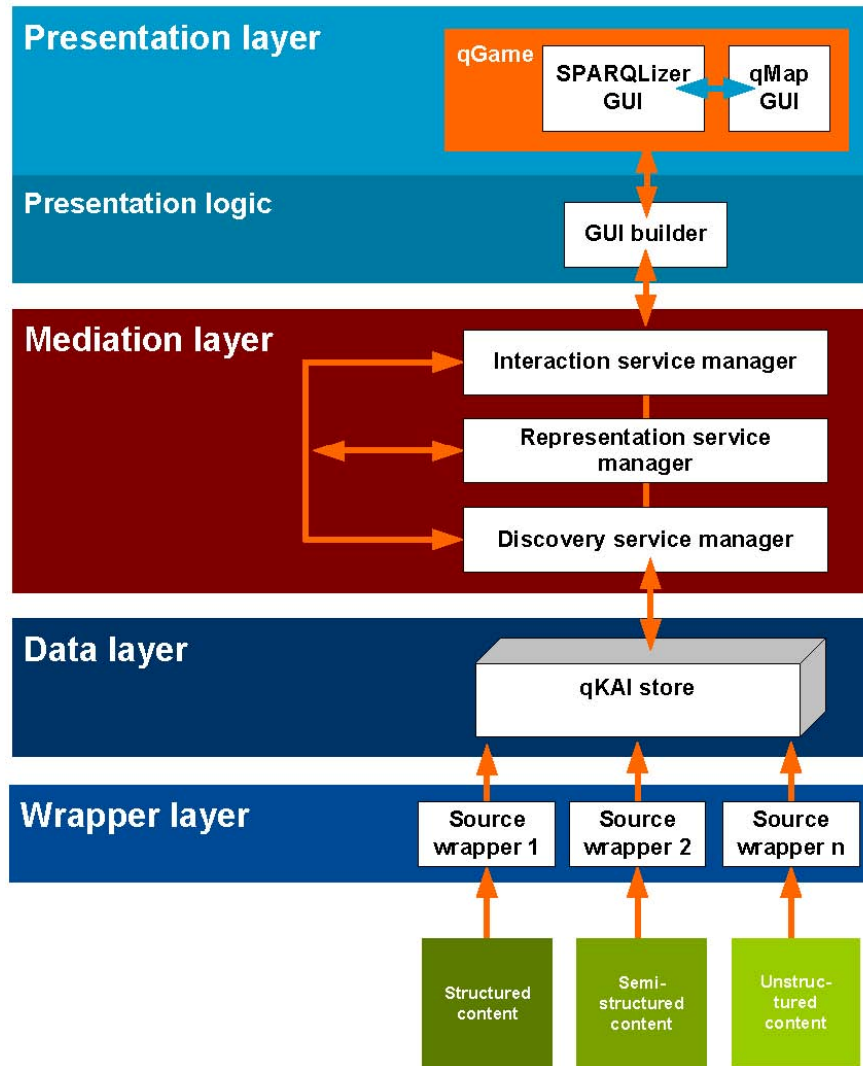
Motivation and requirements

- **How to reuse Open Data targeted in higher-layered applications?**
- Provide standard tasks of knowledge engineering:
 - Acquisition, formalization, representation, visualization.
- Determine and enhance quality of content:
 - Analyzes and enrichment of meta information.
 - User's opinion and knowledge to annotate, rate and rank content.
 - **The more we know about a resource, the better we can reuse it.**
- Tackle extended interaction and incentive for user's attendance:
 - Motivating, enjoyable scenarios inaugurating knowledge games,
 - uniting fun and learning is a proved concept (serious games, assessment),
 - strong focus on usability and lightweight, intuitive interoperation.
- Interoperation as possibility to deduce new knowledge,
- AFAWK: No applications available so far that are based on Open Data for knowledge transfer and learning.

qKAI concept and focus

- We want to offer scenarios based on user-oriented web services for rich and lightweight interaction with Open Data:
 - Therefore we have to implement standard tasks of knowledge engineering for extended interaction as a generic application framework,
 - available Java APIs for subtasks have to be combined and respectively extended,
 - with a scalable, reusable and unifying software concept: Social Semantic Web and **Service Oriented Architecture** principles are considered as Dreamteam, retaining application autonomy with loosely coupling of remote resources and services becomes possible,
 - paradigms like extensibility, addressability, stateless communication are fulfilled.
- Build a hybrid (Meta) data repository for distributed resources:
 - to reach non redundant enrichment of existing resources through semantic interlinking.
- Qualify Open Data through meta data and interoperation:
 - automated analyzes of meta data using the Aperture Java framework,
 - at the user side we enrich content through interaction based on gaming joker options.
- Game-based knowledge transfer is instanced to simplify interoperation and enhance user participation.
 - to enable a proposal of suitable use cases, API- and service compositions for deploying Open Data in knowledge transfer and learning.
 - User-centered with „Rich User Experience“ using Flex/Flash because of high design and functionality issues.

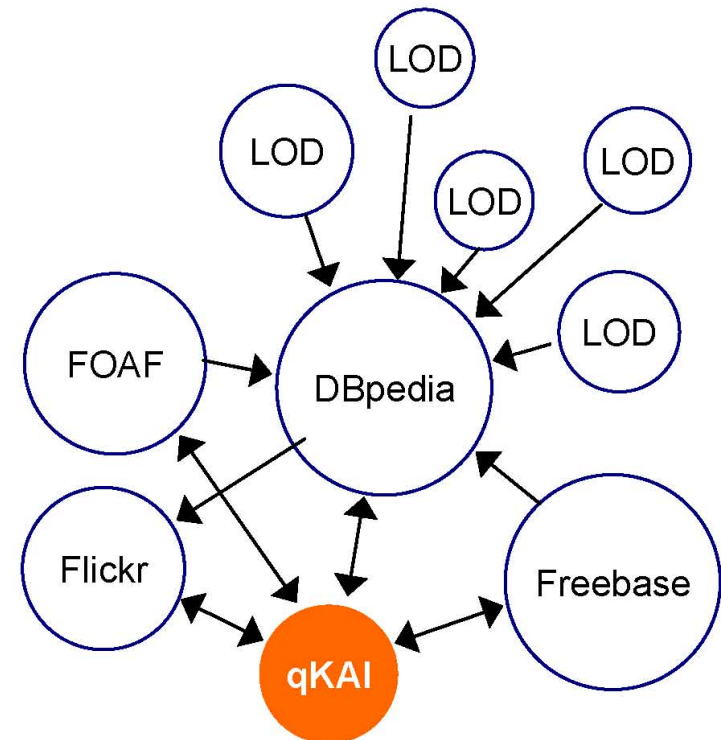
qKAI application framework



- Designed according to a 5-tier-layer design,
- in combination of RIA model, mediator-wrapper concepts and SOA.
- SOA/REST (Representational State transfer) paradigms
 - stateless services and server with communication over http protocol.
- We get a stateful Client using Flex/Flash plugin.
- Desktop-alike applications with advantages like
 - faster reaction to user requests,
 - less network traffic and server load,
 - offline usage possibility.
- We decided for a Rich Thin Client
 - Business logic remains at server side,
 - Rich UI Engine delivers GUIs,
 - the presentation logic is divided from visualization components.
- Mediation layer as business logic or controller.
 - „A mediator is a software module that exploits encoded knowledge about certain sets or subsets of data to create information for a higher layer of applications.“ (Wiederhold, 1992)
 - Middleware, that connects available services with further technical components (Service Mediation).

qKAI data layer: Linked Data

- How to annotate distributed resources with minimal redundancy, access to live resources and ability to change management instead of data dump copies?
- qKAI produces statements about resources
 - for every resource a new qKAI URI is generated,
 - following REST/Linked Data paradigms,
 - we get a Linked Data repository with semantic interlinking using seeAlso or sameAs description in RDF and OWL,
 - qKAI stores only new information and not the provenance source while adding a new node to LOD.
- It is the foundation of a kind of „self-propelled data structure“ that grows by and by with increasing user interaction.



Quality of content

- Meta data can be seen as a quality feature:
 - The more meta data we are snapping, the better we get to know the content.
 - There is no absolute quality, but we can compare resources with each other (Open World Assumption) and weight them based on the amount and structure of meta information.
 - Enrichment of a resource happens in the corresponding qKAI URI file.
- One example is a domain ranking visualized as tag clouds
 - About which domain we get the most information right now?
- First level criteria:
 - Meta data directly included in a resource like format, timeliness, author, provenance, language, ...
 - They can be determined with help of the Java framework Aperture
- Second level criteria:
 - User interaction for example with gaming jokers helps to enhance semantically correctness,
 - after users have played jokers, ranking, annotation and even semantic correction becomes possible.
- Third level criteria:
 - Employing NLP (Natural Language Processing) might detect some more information hidden inside the resource by entity recognition: NER (Named Entity Recognition), PoS (Part of Speech) Tagging or automated summarization, ...

Wissenschaft
Architektur
Kunst&Kultur
Informatik
Geschichte
Geographie

Game-based question answer creation

- We focus on question-answer games first
 - gaming elements are a question, a answer and their assignment as knowledge unit
- **Creating questions and answers is the game itself – so we do not have to create gaming content laboriously before starting to play**
 - Users already get rewarded for creating questions and answers or completing them to a knowledge unit,
- multiple choice, text-text assignment, image-text assignment or ordering questions are available as learning standards in LMS too.
- This question types can be easily converted into IMS/QTI format after in game creation,
- embedding of multimedia is also possible.
- Gaming concepts are described through ontologies and the user makes his choice, what he wants to ask for or what kind of questions he wants to answer
 - type (person, location, object), domain like computer science or architecture, ...), gaming type or question type.
- qKAI concepts use existing ontologies where possible to describe own domains, persons or themes
 - YAGO, SKOS, Dublin Core, FOAF, Dbpedia,
- Game-based rating and ranking of content and interaction is done by in game joker options and a **global point system to reward any kind of interaction.**



Examples of gaming components and sequences

Points of Interest (POI): personalized knowledge view

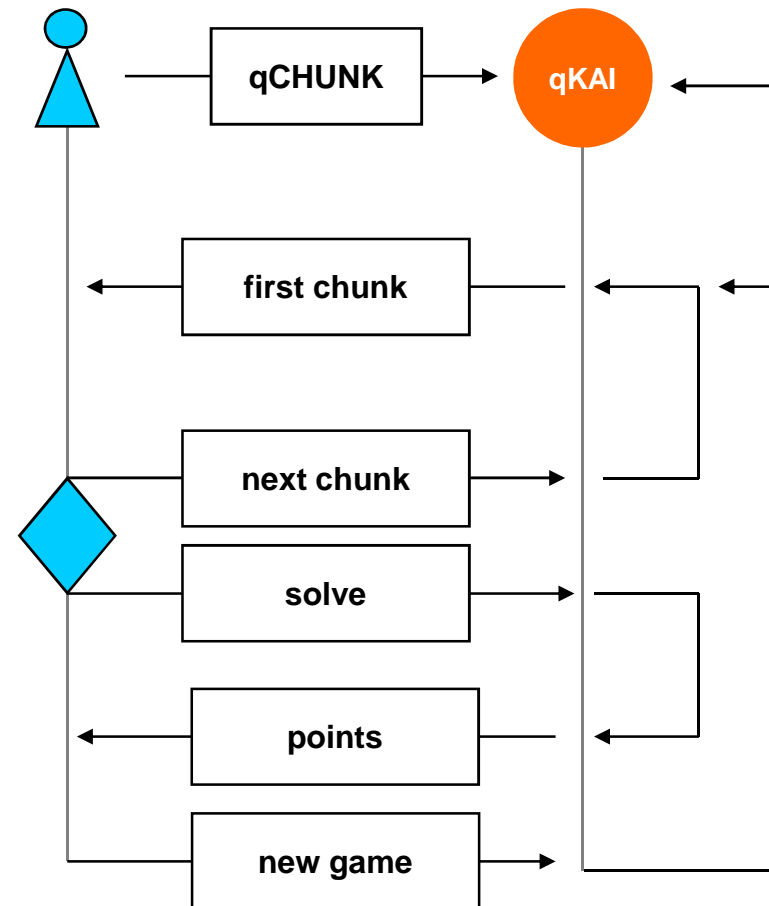
- At the beginning of an knowledge game – or even at many other interaction scenarios - we have to define the domain and the search space the user wants to play or interact with:
 - The web complies to the open world assumption – so we are not able to present a absolute complete domain view.
 - But we can offer what is available and let the user choose out of it.
 - There are lots of domains and themes in the Web of Data.
- POI setter offers a frontend to
 - limit the search space through the qKAI service repository,
 - suggest available content and concepts,
 - filter and select.
 - This can be done thematically, geographically or both.
- qKAI builds a personalized knowledge view this way
 - handled as RDF graphs.

qFOAF: Semantic user profil

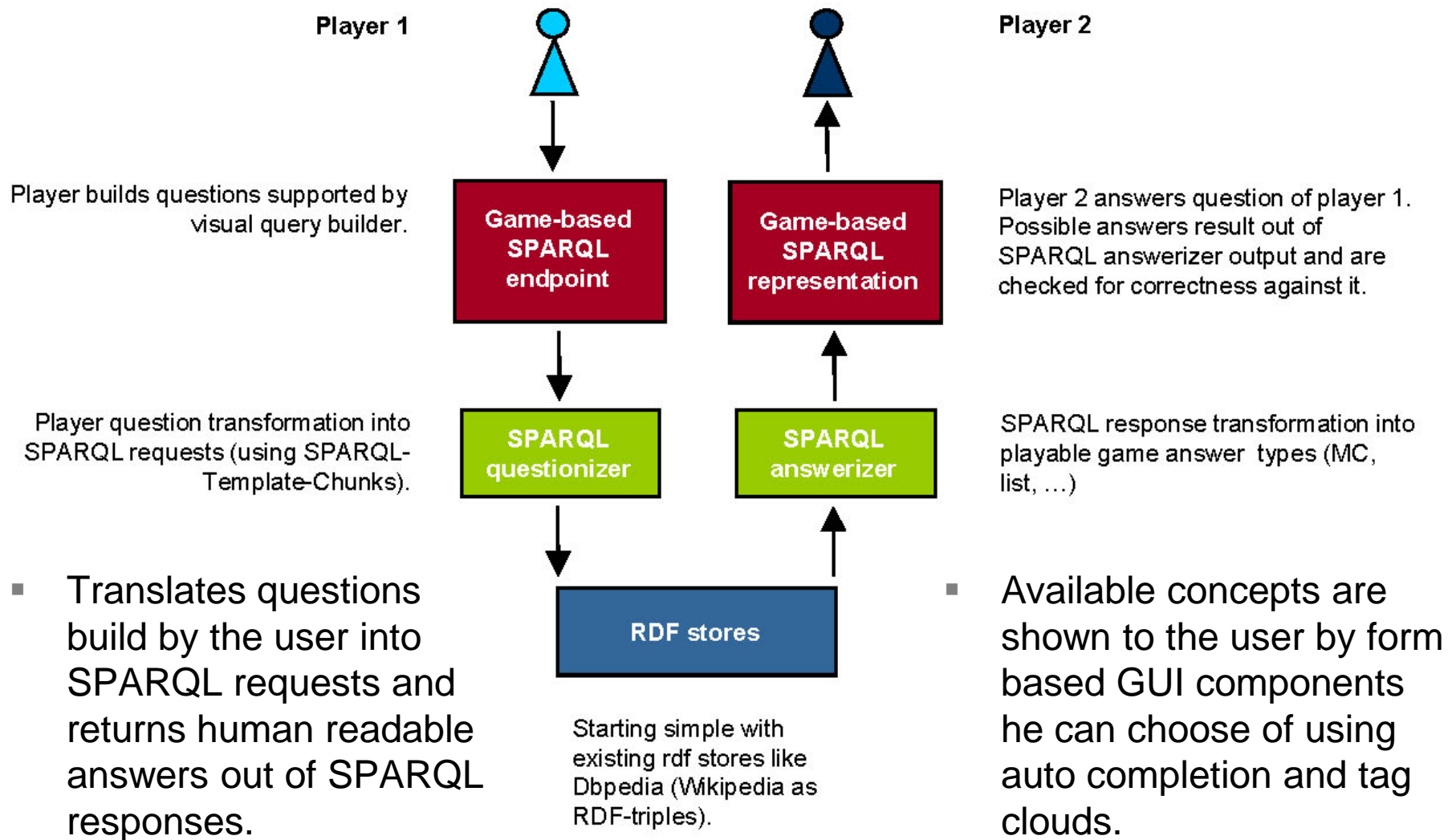
- We need a sustainable user profile to store gaming results and personal information in a transparent manner.
- qFOAF is build as qKAI resource in RDF at the beginning of a game
 - as extended FOAF file (FriendOfaFriend) with unique URI for every user.
 - It connects the user semantically with topics, knowledge units, other players, ... step by step while gaming
 - e.g. <http://qkai.org/foaf/msteinberg/foaf.rdf>
- The qFOAF file can be enriched
 - (semi)automated with given information by the user
 - e.g. by including existing FOAF files, geocodes or interests.
 - while gaming and interacting
 - with game points and content like created questions, answers, knowledge units, ratings, ...
 - with domains, locations, friends, ...

qCHUNK: Game goal and sequence

- The user gets presented small textual information chunks (out of Wikipedia) and has to guess the requested term with as less chunks as possible
- multimedia chunks like zoom parts out of images are conceivable too.
- Chunk: **?? is the capital of Lower Saxony founded in 1942.**
- Answer: **Hanover.**



SPARQLizer: Game-based SPARQL-Endpoint



SPARQLizer: example questions (questionizer)

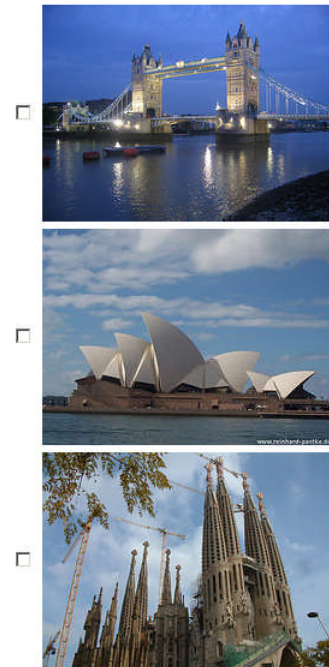
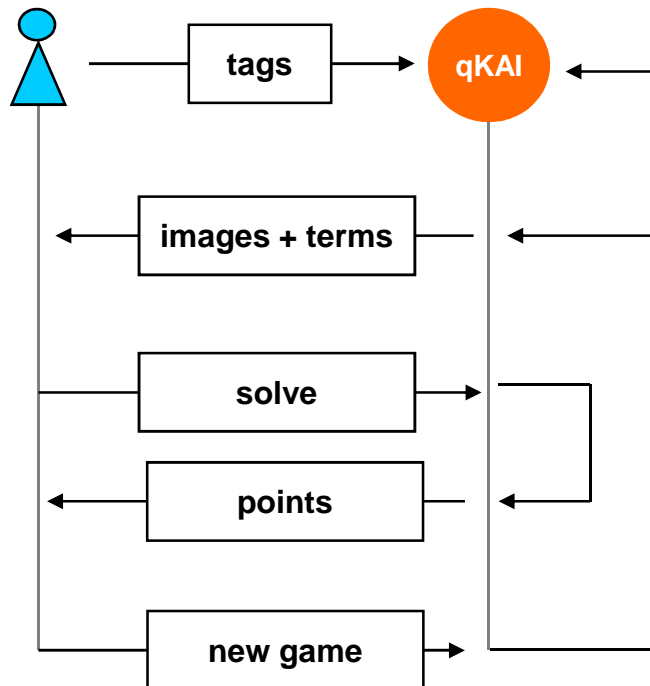
- Which are the capitals of Europe?
- Which cities have famous buildings of the gothic epoch?
- ...
- Which people are born in Berlin before 1900?

```
SELECT ?name ?birth ?death ?person WHERE {  
  ?person dbpedia2:birthPlace <http://dbpedia.org/resource/Berlin> .  
  ?person dbpedia2:birth ?birth .  
  ?person foaf:name ?name .  
  ?person dbpedia2:death ?death  
  FILTER (?birth < "1900-01-01"^^xsd:date) .  
}  
ORDER BY ?name
```

Immanuel Becker
Abraham Mendelssohn Bartholdy
Achim von Arnim
Adalbert von Preußen
...

SPARQLizer: Flickr text image question

- User gets presented randomized terms and images out of Flickr (Web 2.0 image sharing platform) and then has to assign the right term to the right image (multiple choice or drag & drop assignment).



qMAP: Map-based geocode GUI

- Presents a gaming board for available knowledge units (questions and answers)
 - Geocoded placement of locations, events, buildings, photos, persons, ...
 - currently e.g. a photo often knows where it is taken.
- qMAP provides a frontend for interaction with filtering, searching, editing and adding of information or knowledge units.
- OpenStreetMap or Yahoo Maps are good alternatives to Google Maps
 - especially using Flash/Flex API.
- Map symbols are connected with different gaming interactions and information units.



qRANK

- Allows game-based rating and ranking of resources, information and knowledge units by joker options,
- Know-it-all-joker bounds player to add missing information,
- Nonsense-joker marks information as semantically wrong and defers it to review mode by other users,
- Explorer-joker with predefined time, history-joker enables lookups in played transaction protocol, ...

qLINK

- Offers game-based interlinking with new resources,
- Linked Open Data basis will be enhanced,
- new interlinking brings qPOINTS in qFOAF.

qPOINT

- Is a global point and level system documenting learning progress and personal knowledge,
- **every interaction is rewarded with qPOINTS,**
- based on an interaction catalog with assigned points,

Current work and outlook

- **Current:**
 - We are implementing demonstrators für GUI and background functionality:
 - qCHUNK, SPARQLizer, POI-Setter are work in progress.
 - Backend services for qCHUNK and POI-Setter are almost done, Flex GUI frontend is currently under development.
 - Embedding available resources as qKAI knowledge base using Sindice web service and SPARQL is also work in progress.
- **Next:**
 - Implementing SPARQLizer, qMAP, qPOINT, qFOAF,
 - evaluation of finished gaming types with large student groups.
 - Proof of concept.
- 1,5 years left for prototypically implementation and evaluation of gaming components and services.
- **Long term objective:**
 - Atomic, simple and composite services for generic rearrangement in several scenarios.

Conclusion

- Requirements are derived to reuse Open Data for knowledge transfer and learning.
- Shown on the example of knowledge games:
 - Implemented atomic services allow rearrangement and are suitable for any other purpose in information and knowledge transfer – not only for game-based use cases.
 - Services are adaptable to several domains like computer science, city guiding, history or architecture studies, philosophy, music theory, ...
- Summarizing the main aspects in qKAI:
 - We are designing and implementing a framework specification following REST, Linked Data und RIA paradigms,
 - building a hybrid data store for distributed web resources with semantic interlinking,
 - creating incentive game-based interoperation using scalable web services and rich interoperation,
 - determining and enhancing contents' quality by analyzing meta data and enriching content with user's in game annotations.
- Reusing and composing is a precept at all levels, the challenge is to merge and to expand:
 - Resources, concepts (ontologies), frameworks, APIs, tools, libraries, ...
 - We are aiming at standardized, machine- and human readable staging of Open Data with lightweight interoperation.

Thank you very much for your attention!