Developing a machine-actionable process model for ethical approval workflows

Verena Schwaiger

Master Thesis Presentation

Supervisors: Prof. Dr. Stefan Decker, PD Dr. rer. nat. Cord Spreckelsen
Advisors: Dr. Oya Deniz Beyan, Daniel Mietchen, Ph.D.

AGENDA

1. Introduction

2. Analysis

3. A formal model for ethical approval workflows
   1. Ontology design
   2. Pattern language

4. Concept implementation

5. Case study

6. Summary & Outlook

7. Demo

INTRODUCTION

- Ethics approval process
  - Studies require approval from ethics boards (depending on subject)

- Reuse and discovery of past ethics approval-related documents
  - No open databases

- Openly available approval decisions and process components
  - Transparency: Why did the study need ethics approval?

- Data management integration

BACKGROUND

- FAIRness
  - Make data findable, accessible, interoperable and reusable

- Machine-actionable data management plans

1 Integrate DMPs with the workflows of all stakeholders in the research data ecosystem
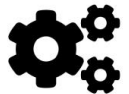
2 Allow automated systems to act on behalf of stakeholders

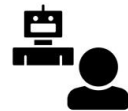3 Make policies (also) for machines, not just for people

4 Describe—for both machines and humans—the components of the data management ecosystem

5 Use PIDs and controlled vocabularies

6 Follow a common data model for maDMPs

7 Make DMPs available for human and machine consumption

8 Support data management evaluation and monitoring

9 Make DMPs updatable, living, versioned documents

10 Make DMPs publicly available

Source: Miksa, Tomasz et al: *Ten simple rules for machine-actionable data management plans* (2018)

PREVIOUS WORK

- For machine-actionable DMPs: DMPTool, DMPRoadmap

- deon: ethics checklist locally saved
  - for data science in general

- No existing applications for ethics approval processes

https://dmptool.org/

https://github.com/DMPRoadmap/

http://deon.drivendata.org/

ANALYSIS

- Analysis of ethics approval guidelines
  - \> 20 universities, other institutions
  - Similar processes
  - Application either digital or on paper, stored internally
- Experience reports
  - Observation of an ethics committee (Martín-Arribas 2012)
    - → Lack of ethics education, administrative errors
  - Hospital study (Watson, Rayner, Lumley 2007)
    - → Difficulty communicating and coordinating with 87 hospitals
  - University students study (Vadeboncoeur, Townsend, Foster)
   - → Pre-existing studies
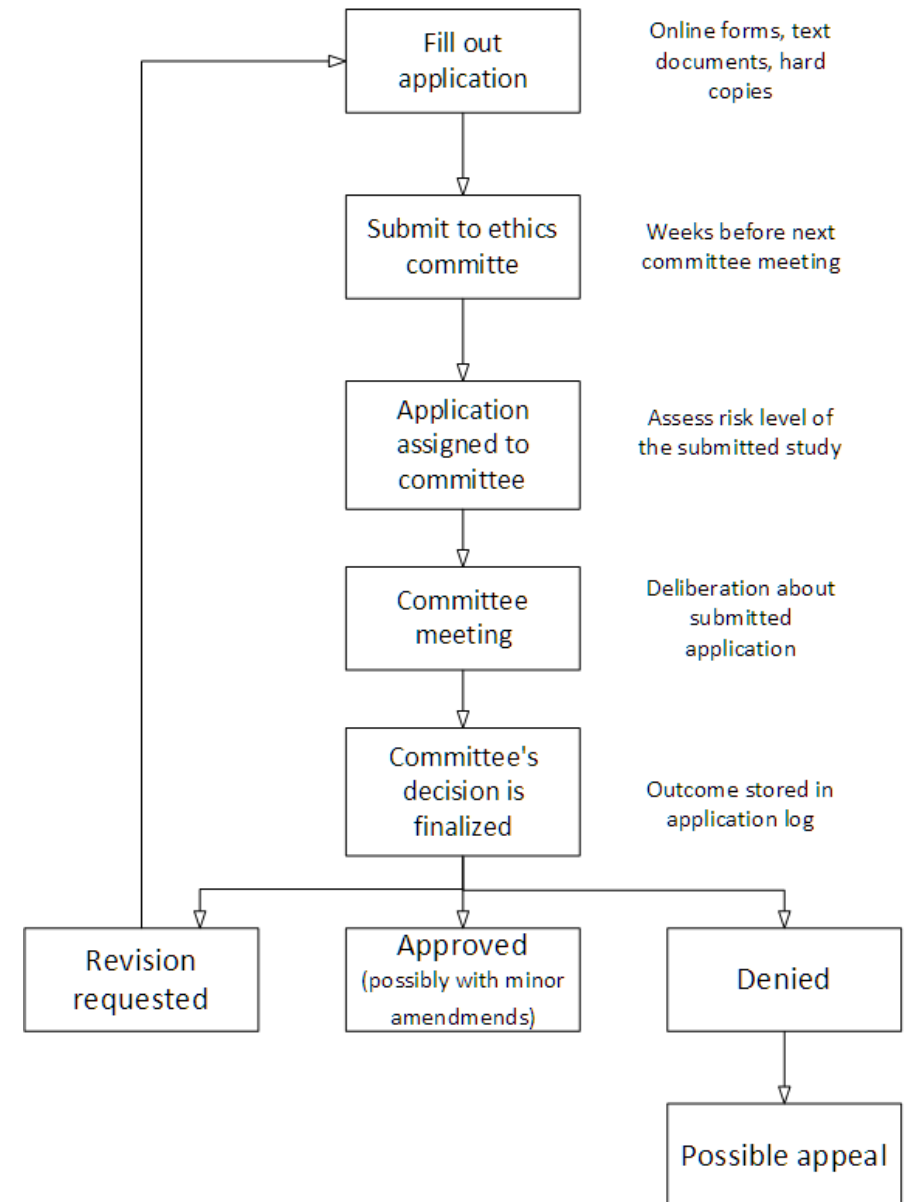
ANALYSIS

- Special cases

- Different requirements and ethical issues depending on subject

- Requring ethics approval not always obvious
  - Easy: humans, children, …
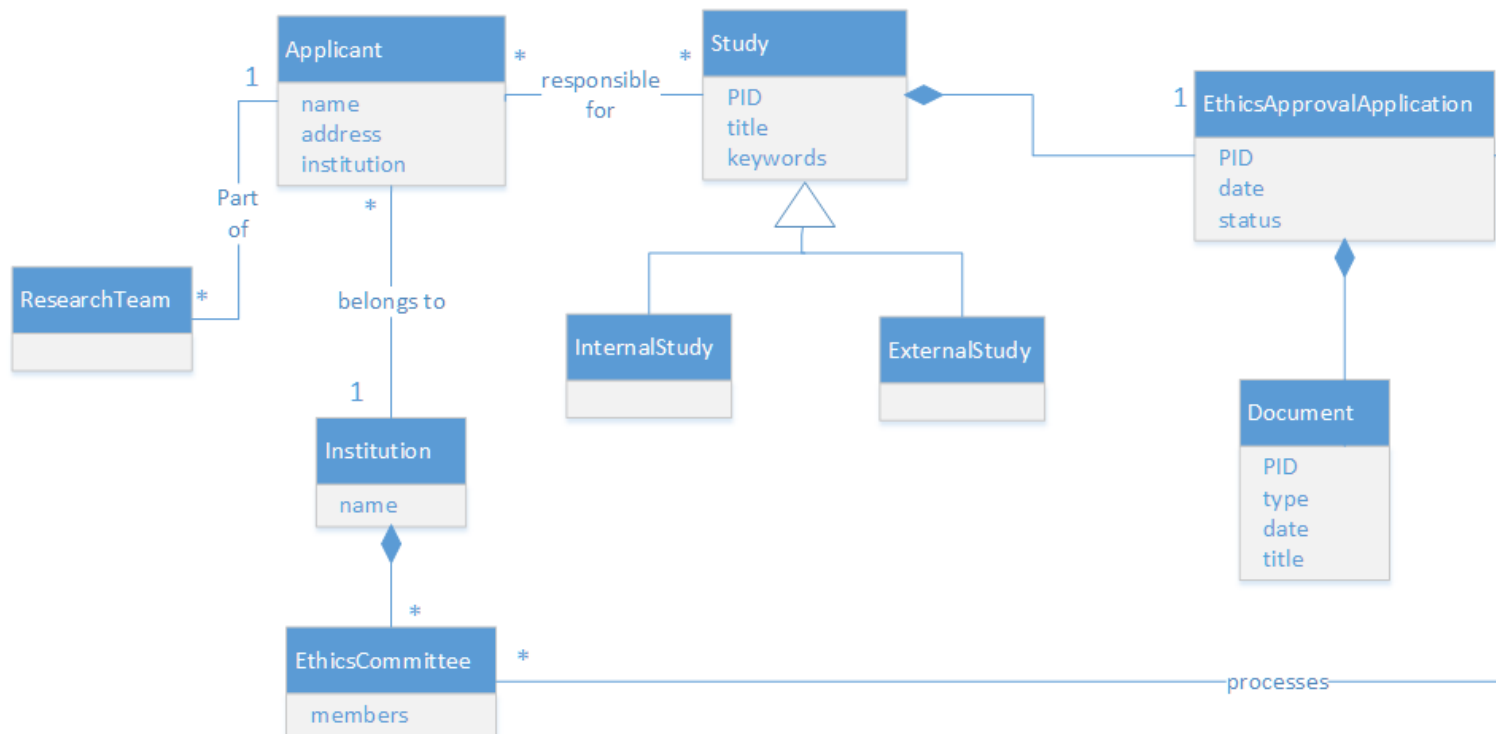 - Harder: potential misuse of results, questionable funding, …
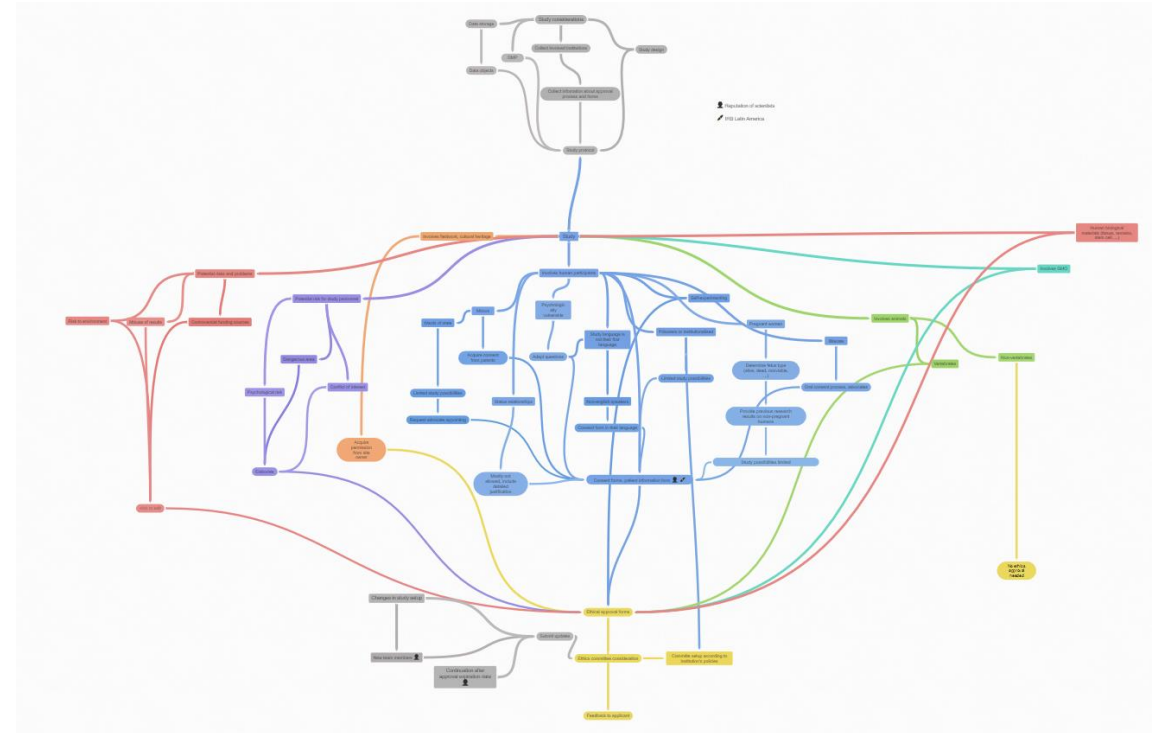
ANALYSIS

- Special cases

- Different requirements and ethical issues depending on subject

- Requring ethics approval not always obvious
  - Easy: humans, children, …
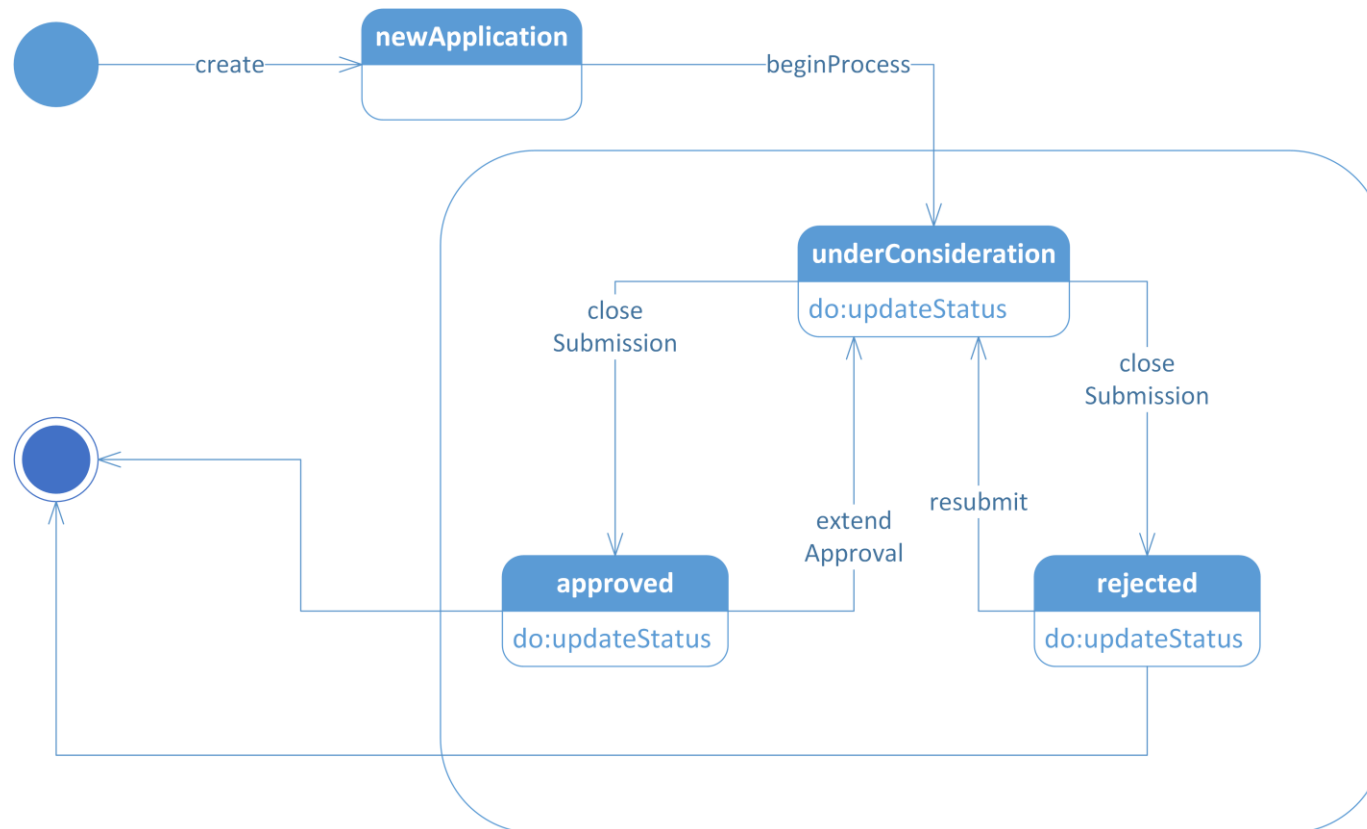 - Harder: potential misuse of results, questionable funding, …
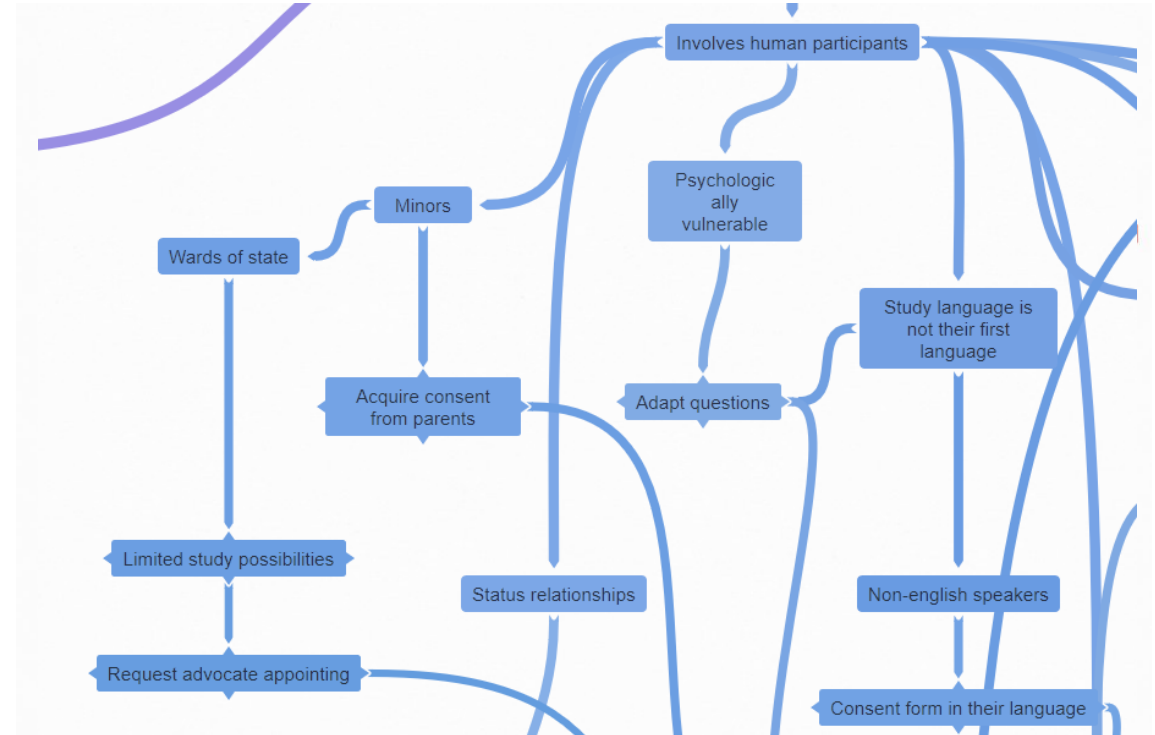
FORMAL MODEL – PATTERN LANGUAGE
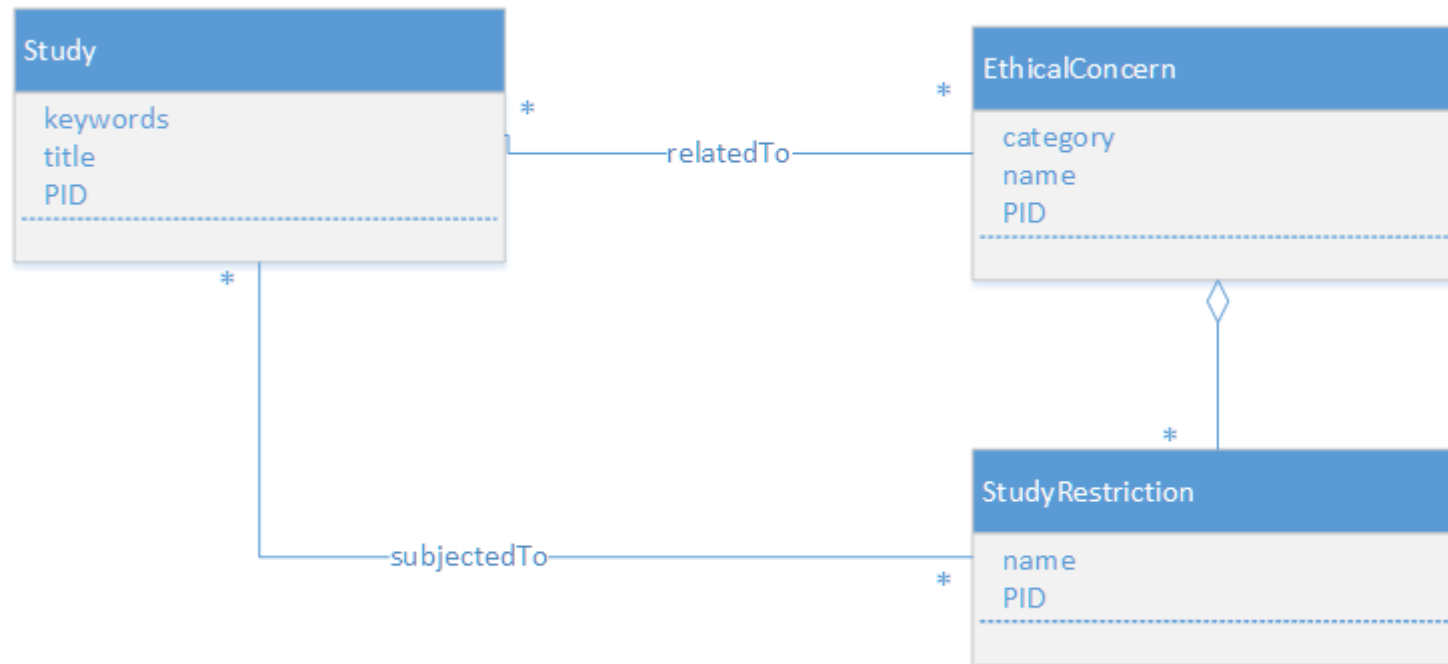
1. Describing ethics approval process instances

FORMAL MODEL – PATTERN LANGUAGE
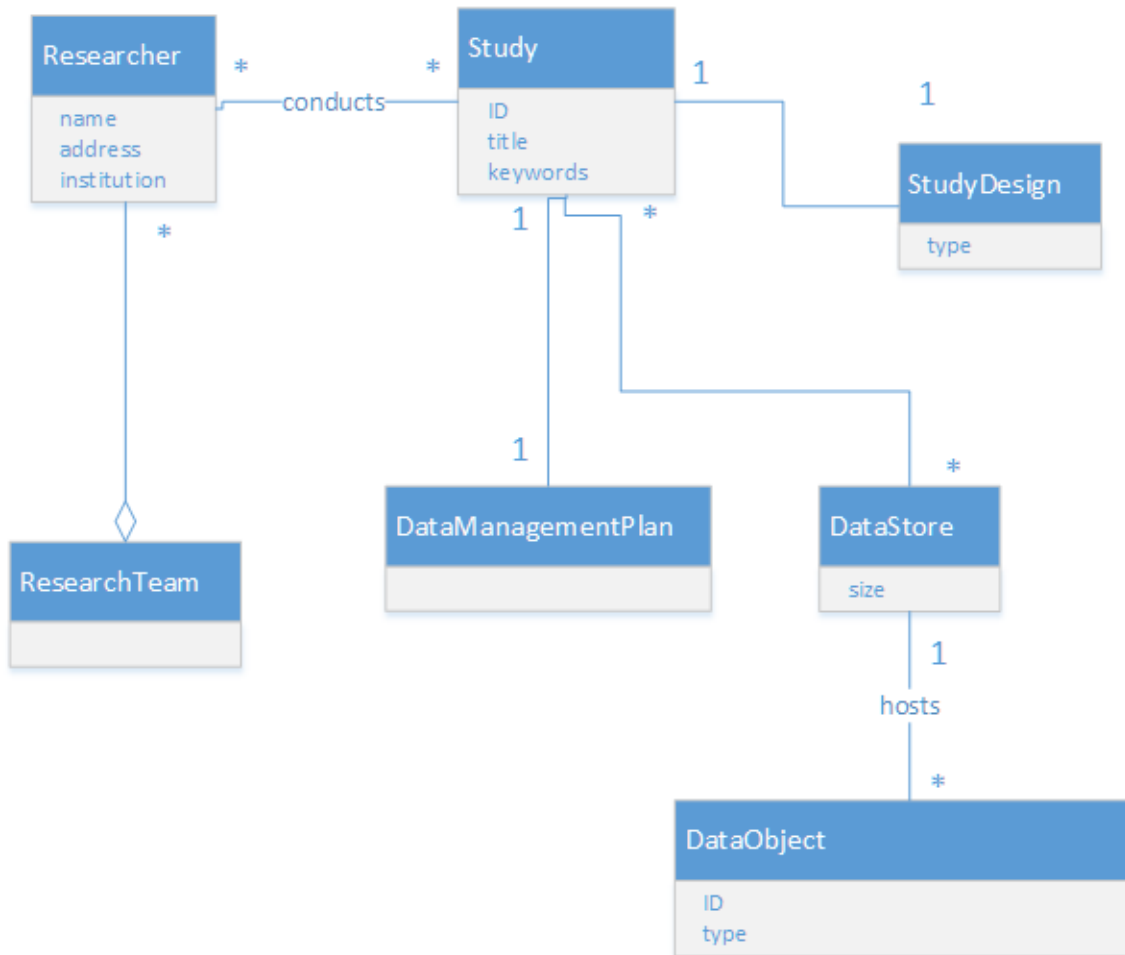
1. Describing ethics approval process instances

FORMAL MODEL – PATTERN LANGUAGE

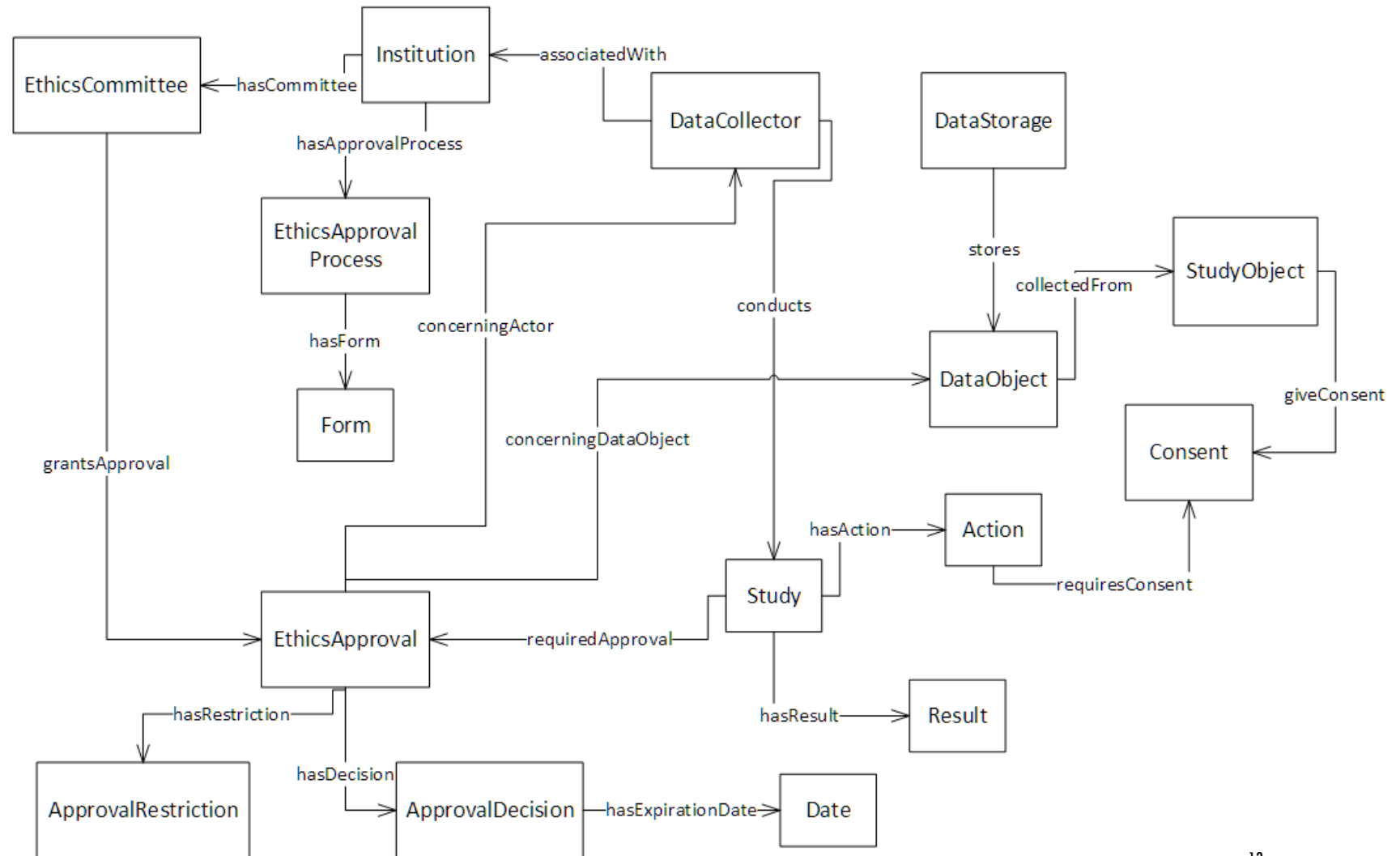2. Describing components that could influence specific workflows

FORMAL MODEL – PATTERN LANGUAGE

3. Describing the study itself

FORMAL MODEL – ONTOLOGY

- Integration of existing ontologies (DUO, ICO, …)

- Controlled vocabulary for later use

CONCEPT IMPLEMENTATION

1. Potential features for stakeholders

2. System design

3. Ethics data storage

4. Evaluation

5. Case study

FEATURES FOR STAKEHOLDERS

Study author perspective	Feasibility
Identify application requirements	◯
Checklist (required documents)	✓
Managing application data	✓
Show similar studies	✓
Public file storage	✓
Timeline	✗
Predicting approval/rejection	✗

FEATURES FOR STAKEHOLDERS

Research perspective	Feasibility
Browse existing application data	✔
Specific queries	✔
Access application documents	✔

Patient perspective	Feasibility
Access background information on studies	✔

FEATURES FOR STAKEHOLDERS

Ethics committee perspective	Feasibility
Access decisions of other institutions	✓
Reuse publicly available documents	✓

SYSTEM DESIGN
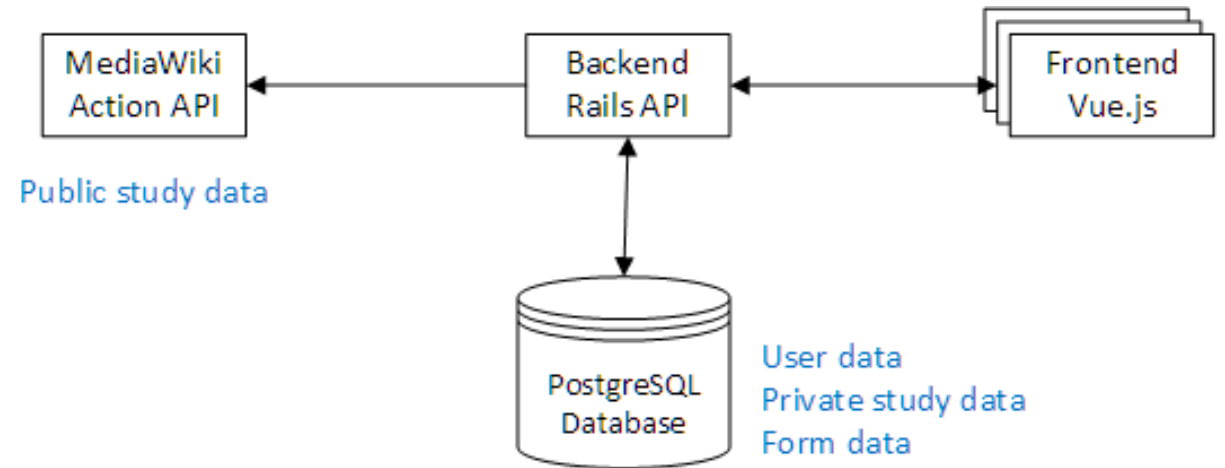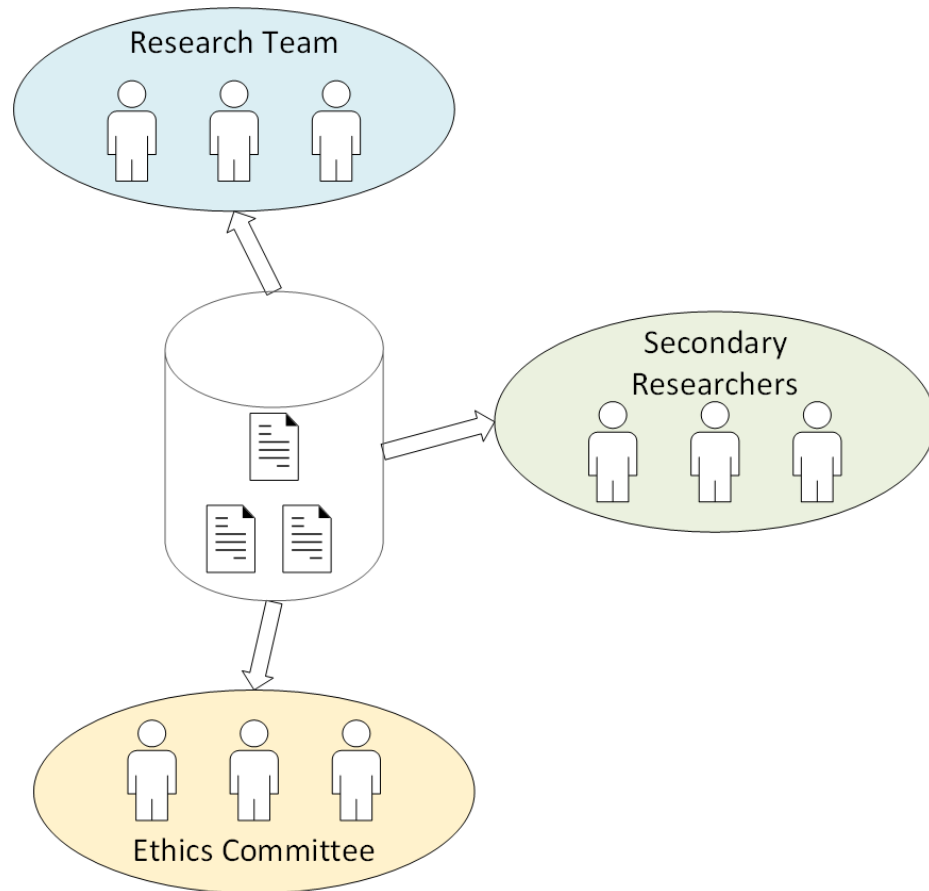
- Potential integration with existing DMPTool
- Backend: Ruby on Rails API application with PostgreSQL database
- Frontend: Vue.js for reactive interface

SYSTEM DESIGN

- Data <> Metadata

- Data stores
  - Public: files, study information, … → focus on researching and querying study data
  - Private: study information, user data → focus on managing approval applications

MANAGING USER WORKFLOWS

1. Identifying application requirements

2. Managing application data

3. Checklist

4. Similar studies

MANAGING USER WORKFLOWS

Identifying application requirements

- Option 1: Automatically through natural language processing
  - Easy for the user
  - Unreliable, not suitable for complex studies

- Option 2: User questionnaire
  - Questions based on collected application information
  - More transparent
  - Precise

MANAGING USER WORKFLOWS

Organizing questionnaire data

- Suitable format for storing questions

- JSON, XML, RDF ?

- XML: tree structure, not ideal for reading and extracting data

- JSON: tree structure, easy processing for web applications

- RDF: graph structure

➡️ JSON works best for storing questions for the web application

MANAGING USER WORKFLOWS

Identifying application requirements

Home   Studies   Add Study

Study title [   ]

Study details

☐ Are human participants involved?

☐ Does your study involve human or other biological tissue?

☐ Does your study involve genetically modified organisms?

Assessing walking behavior on small town sidewalks

Edit/Create Wiki page   Checklist   **Update/Review Questionnaire**   Similar Studies

☑ Are human participants involved?

Informed consent is required

☑ Are minors involved?

Parental consent is required

☐ Are any involved minors wards of the state?

☐ Are pregnant women involved?

☐ Are prisoners involved?

☑ Are participants involved who aren't native speakers of your study's language?

MANAGING USER WORKFLOWS

Managing application data

- User input: study title, keywords, questionnaire data, …

- Identify potential ethical issues through questionnaire input

- Users may want to store input privately first

- Solution
  - PostgreSQL database with the Rails API backend
  - CRUD operations through axios requests from frontend

MANAGING USER WORKFLOWS

Managing application data

MANAGING USER WORKFLOWS

Checklist

- Based on answered questionnaire

- Show required documents (consent form, …)

- Link to file upload for the public storage

Similar studies

- Show similar studies based on input data

- Simple topic-based recommendations

- Open for future improvements: semantic similarity, language-independence, …

Assessing walking behavior on small town sidewalks

Edit/Create Wiki page   Checklist   Update/Review Questionnaire   Similar Studies

☑ Consent form
☐ Patient information sheet (Upload to Wiki)
☑ Parental consent form
☐ Acknowledgement form for observation and data collection (Upload to Wiki)

Save

Assessing walking behavior on small town sidewalks

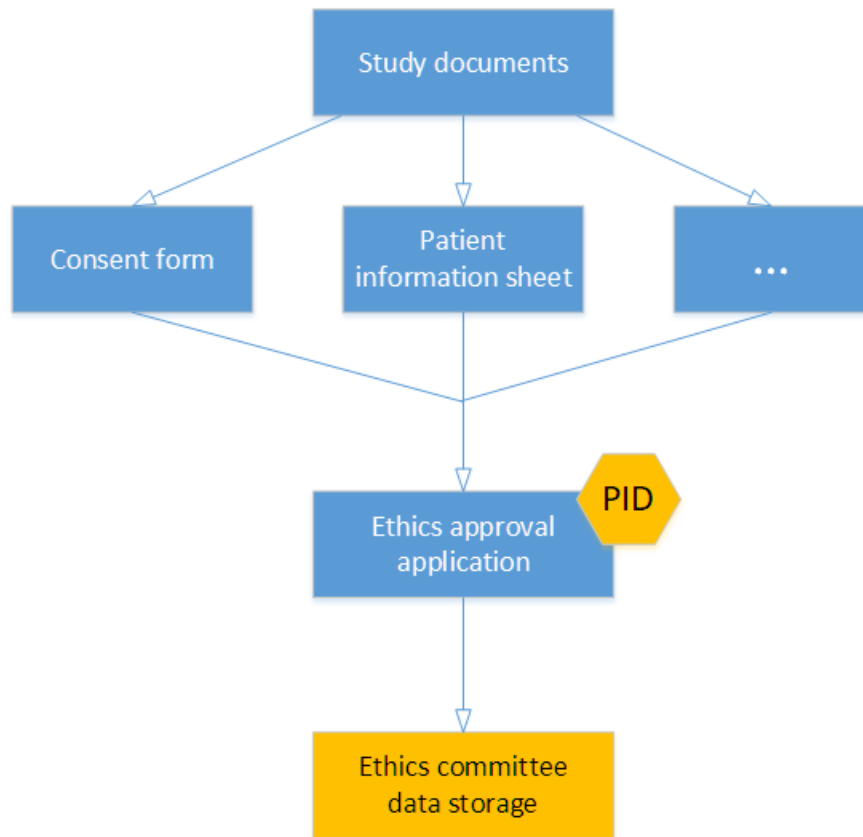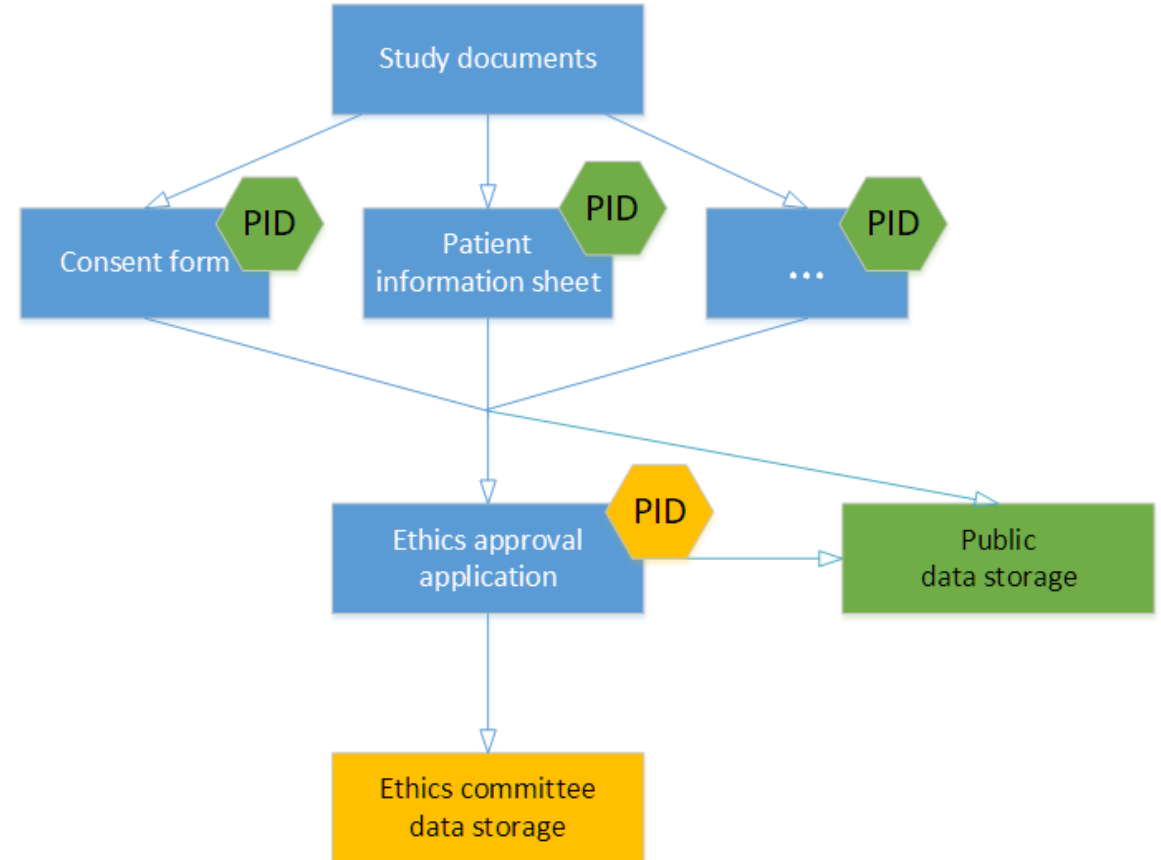Edit/Create Wiki page   Checklist   Update/Review Questionnaire   Similar Studies

Predicting walking routes on weekdays

PIDS

Current state
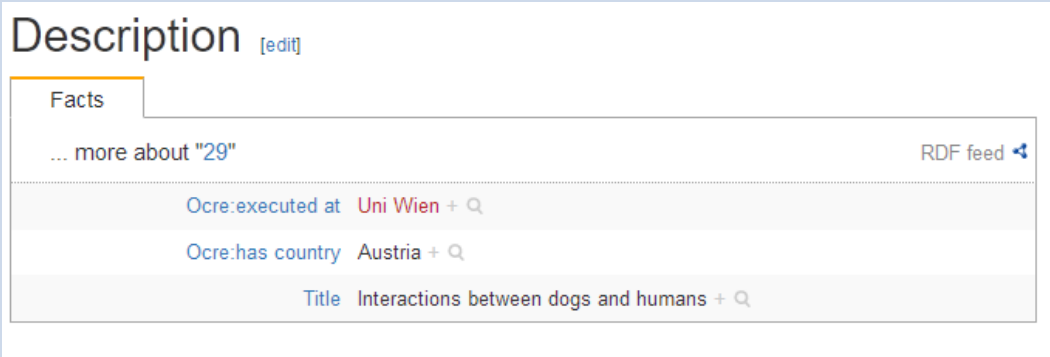
Improvement

MEDIAWIKI FOR PUBLIC STORAGE AND PIDS

- Public storage requirements
  - Querying existing data
  - Browsing studies
  - Semantic links
  - PID support
  - Versioning

- MediaWiki
  - Users are familiar with Wikis
  - Adaptability through extensions

MEDIAWIKI SOLUTIONS

Wikibase	Semantic MediaWiki
RDF store possible → SPARQL queries	
Ontology import	
Data stored as statements	Semantic annotations
Made for Wikidata → less adaptability	Higher adaptability with extensions

GENERATING PIDS

1. Study PID generated upon creation in the system

2. File PIDs generated during file upload

75 - 8726 PID

Study ID   Random number
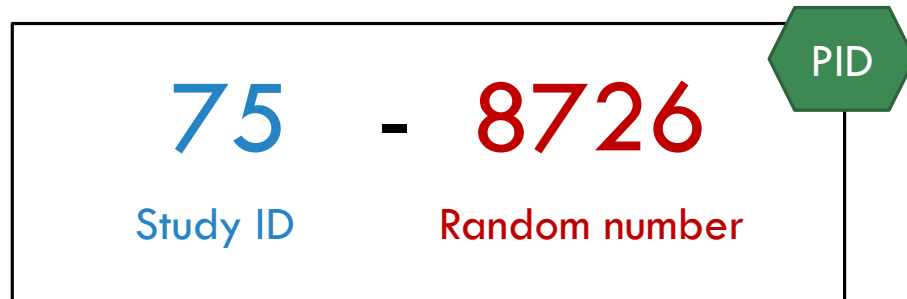
3. URI with study ID as PID: http://mediawiki/75

4. PIDs associated with files and studies on MediaWiki

→ Searchable, versioned

→ After deletion not findable → Future: Metadata?

EVALUATION – PRINCIPLES FOR MADMPS

- Implementation not yet for all stakeholders (e.g. ethics committees)

- Successful use of controlled vocabularies and PIDs

- RDF triplets organize data for machine consumption

- Public storage on MediaWiki for human consumption and monitoring

- Updatable through web application, MediaWiki provides versioning

- No automated acting on behalf of stakeholders → Integration with DMPTool

CASE STUDY

Source:

Ral Madrid. The rise of ethnic politics in Latin America, 2016

https://data.qdr.syr.edu/dataset.xhtml?persistentId=doi:10.5064/F6MS3QNV

Home   Studies   Add Study

Study title The rise of ethnic politics in L

Study details

☑ Are human participants involved?

Informed consent is required

☐ Are minors involved?

☐ Are pregnant women involved?

☐ Are prisoners involved?

☐ Are participants involved who aren't native speakers of your study's language?

☐ Are any illiterate participants involved?

☐ Are psychologically vulnerable participants involved?

☐ Are any of the researchers a study subject themselves?

☐ Is there any risk of psychological or physical harm for participants?

☐ Are there any potential gatekeepers, such as schools or nursing homes?

CASE STUDY

The rise of ethnic politics in Latin America

Edit/Create Wiki page   **Checklist**   Update/Review Questionnaire   Similar Studies

☐ Consent form (Upload to Wiki)

☐ Patient information sheet (Upload to Wiki)

☐ Acknowledgement form for observation and data collection (Upload to Wiki)

Save

CASE STUDY

Files [edit]

File	Description	PID
0 Madrid IRB short consent form (Spanish 2006).pdf	Consent form	87-9

Description [edit]

Project Summary: This research, which was eventually published in a 2012 book by Cambridge University Press entitled The Rise of Ethnic Politics in Latin America, focused on the emergence of indigenous parties in Latin America. Specifically, it sought to explain why some parties based in the indigenous population succeeded while others failed. The study focused on the three South American countries with the largest indigenous populations--Bolivia, Ecuador, and Peru--but a comparative chapter examined the fate of indigenous parties in the rest of Latin America as well. The central argument of this study is that indigenous-based parties have succeeded in recent years by using inclusive ethnic and populist appeals to reach out to whites and mestizo as well as indigenous people. Indigenous parties, unlike many other ethnic parties, have managed to win support across ethnic lines because the long history of racial mixing in Latin America blurred ethnic boundaries and reduced ethnic polarization.

Data Abstract: This study used a combination of qualitative and quantitative data, including interviews, party documents, journalistic accounts, surveys of public opinion and municipal-level census and electoral data. The data consist of notes in Spanish from interviews with prominent party leaders, legislators, interest group representatives, government officials, and pollsters. I selected interviewees who were deemed to have extensive knowledge of the elections and the parties involved in them and the interest groups that supported them. I was particularly interested in interviewees who were knowledgeable about or involved with the Movimiento al Socialismo (MAS) in Bolivia, Pachakutik in Ecuador, the Partido Nacionalista Peruano (PNP) in Peru, and Winaq in Guatemala. The interviews were conducted in 10 summer research trips to Latin America between 2002 and 2008. The interviews were unstructured in nature and were conducted by the author.

Facts

... more about "87"   RDF feed ◄

Duo:is about	ethnic parties + 🔍 , indigenous movements + 🔍 and indigenous parties + 🔍
Maeo:hasDecision	Approved + 🔍
Maeo:hasEthicalConcern	human + 🔍 and HumanParticipant + 🔍
Maeo:hasPID	2004-06-0026 + 🔍
Ocre:executed at	University of Texas at Austin + 🔍
Ocre:has actor	Raúl Madrid + 🔍
Ocre:has country	United States of America + 🔍
Title	The rise of ethnic politics in Latin America + 🔍

CASE STUDY – EVALUATED

- Original study only has consent form

- Requirements not uniform across institutions

- Original uploaded study data has private documents

  → Metadata could be introduced for private data
  → Different requirements are harder to solve

SUMMARY

- Problems:
  - Reuse and discover of ethics approval data
  - Administrative difficulties with ethics approval processes

- Machine-actionable data management plans

- Formalization of the ethics approval process
  - Controlled vocabulary, pattern language

- Concept implementation
  - For stakeholders: study authors, researchers, patients
  - Public Semantic MediaWiki, web application for study authors

OUTLOOK

- Integration of other stakeholder workflows (ethics committees)

- Willingness to open up the ethics approval process

- Possibilities for enhancement with more real-life application data

- Meta-data for ethics approval applications

ANY QUESTIONS?

Thank you for your attention!

Presentation DOI:
10.5281/zenodo.3251763