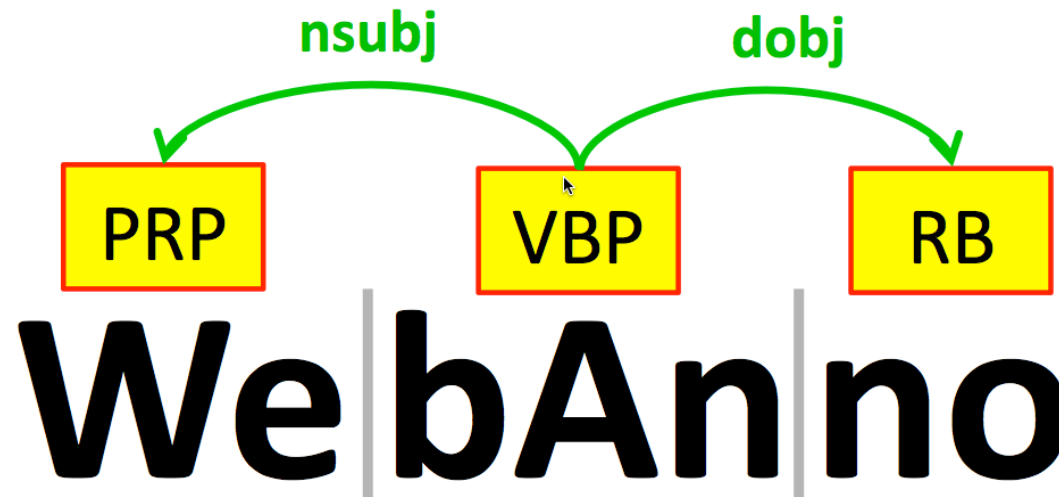


WebAnno: a flexible, web-based annotation tool for CLARIN



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Richard Eckart de Castilho, Chris Biemann,
Iryna Gurevych, Seid Muhie Yimam



#WebAnno



UBIQUITOUS
KNOWLEDGE
PROCESSING



This work is licensed under a [Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/).
If you are interested in using this material under different conditions, please contact us.



WebAnno – an annotation tool for text



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- **Team tool**
 - Allows a distributed team of annotators to work on a corpus
 - Supports different roles within the team (e.g. user / manager)
- **Flexible**
 - Multi-layer annotation with configurable annotation layers
 - Different annotation modes including correction and learning modes
- **Web-based**
 - Available to annotators everywhere, no installation effort
 - All configuration performed through the web interface
- **Platform independent**
 - Platform independent Java-based application
- **Open source**
 - Allows the community to participate

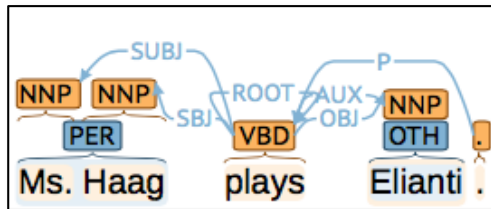
WebAnno – an annotation tool for CLARIN



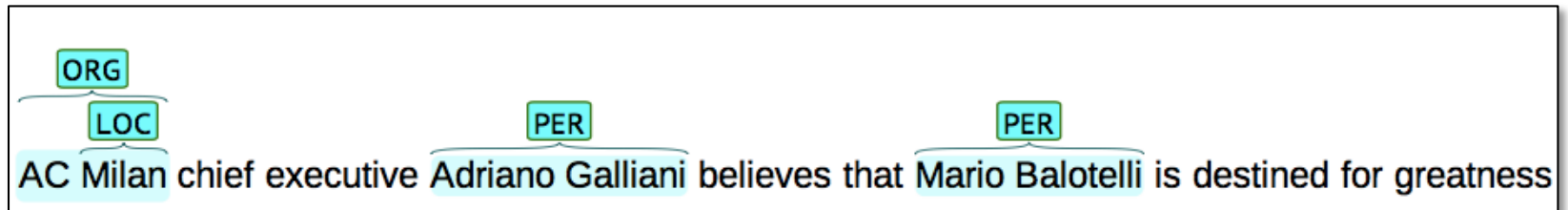
- **Developed based on the requirements of CLARIN F-AG 7...**
 - Dipper et al. NoSta-D: A corpus of German non-standard varieties. *Non-Standard Data Sources in Corpus-Based Research (2013): 69-76.*
 - Benikova et al. NoSta-D Named Entity Annotation for German: Guidelines and Dataset. *Proceedings of LREC. 2014.*
- **... but also used beyond F-AG 7**
 - Pedersen et al. Semantic Annotation of the Danish CLARIN Reference Corpus. *Proceedings 10th Joint ISO-ACL SIGSEM Workshop on Interoperable Sem. Annotation. 2014.*
- **... used and recognized beyond CLARIN**
 - Search “WebAnno” on Google Scholar
 - See our public users mailing list
- **WebAnno is the first annotation tool to supporting WebLicht TCF**
 - Worked with TCF developers to improve TCF support updating files!
- **WebAnno team is constantly in touch with the community**
 - Visit <http://webanno.googlecode.com> after the talk to participate in our survey!

Annotation examples

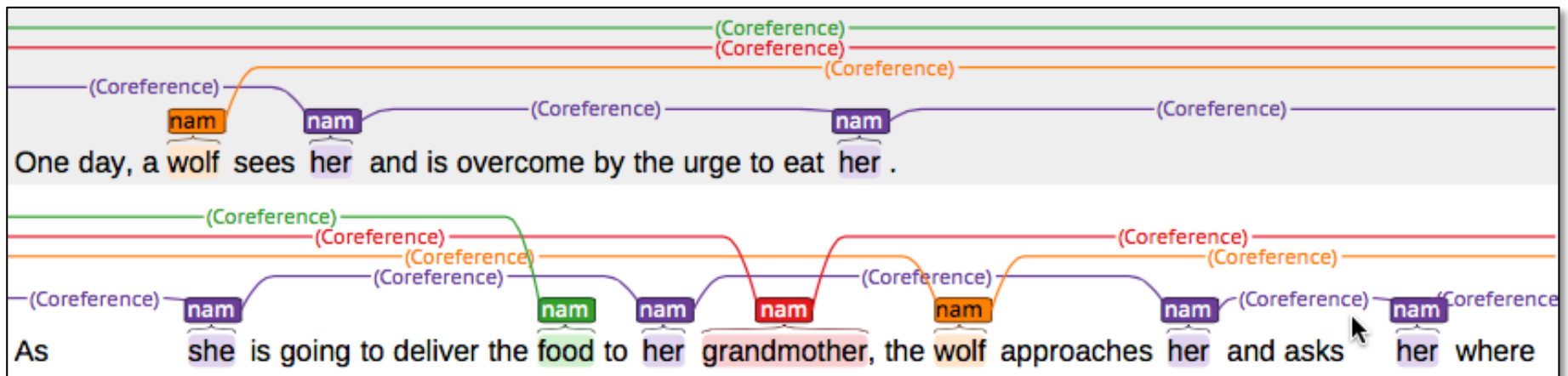
Part-of-Speech & syntactic dependencies



Named entities



Co-reference



Main Menu



[Annotation](#)

- Annotate texts from scratch



[Correction](#)

- Review and correct previously annotated documents



[Automation](#)

- Employ integrated machine learning capabilities



[Curation](#)

- Compare annotations from different annotators and merge them



[Monitoring](#)

- Assign workload to annotators and monitor their progress



[Projects](#)

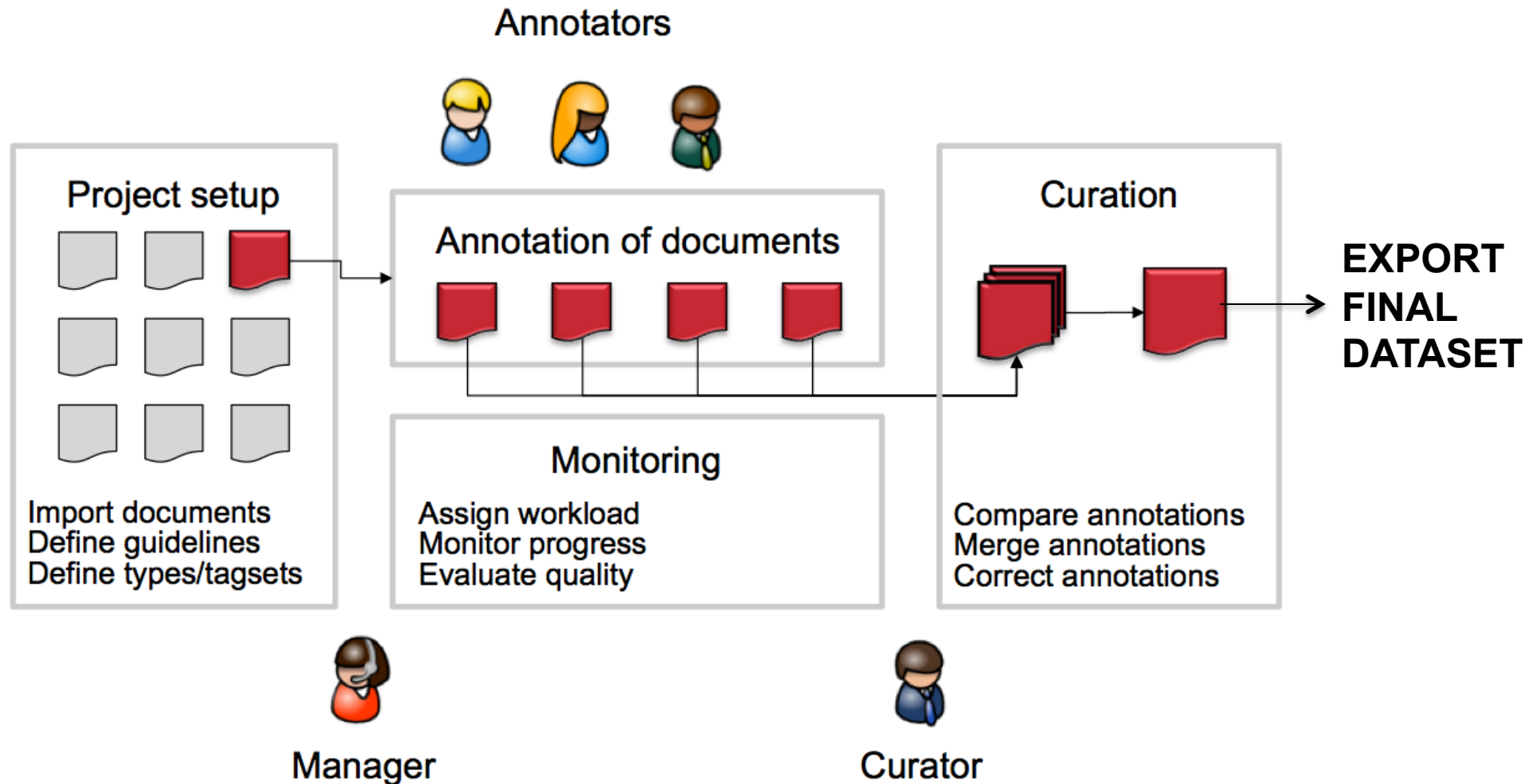
- Create new projects



[Manage users](#)

- Create new user accounts

Workflow of a WebAnno project



Curation



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Curation
WebAnno | Home | User: guest | Log out CLARIN-D

Document | **Page** | **Help** | **Workflow**

Open | Re-create Merge | Prev. | Next | Export | Settings | First | Prev. | Go to | Next | Last | Guidelines | Done

demo-anno-en/PosDep.tcf

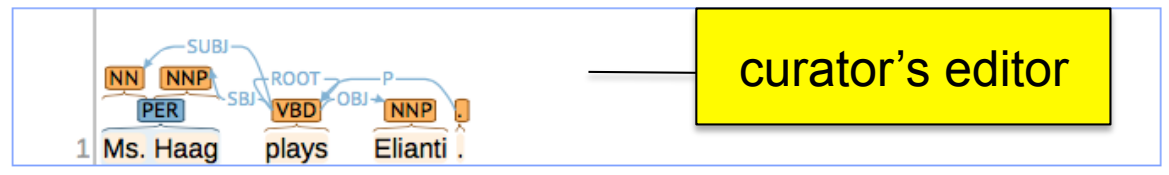
showing 1-5 of 7 sentences

Sentences

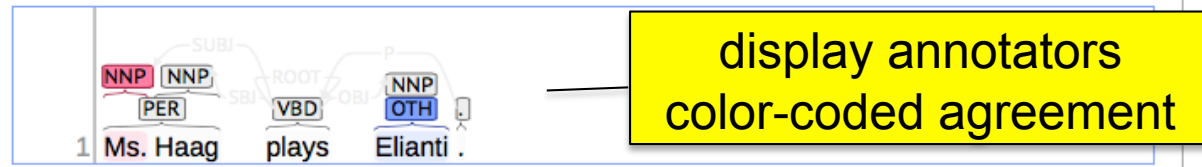
- 1 Ms. Haag plays Elianti .
- Rolls-Royce Motor Cars Inc. said it
- 2 expects its U.S. sales to remain steady at about 1,200 cars in 1990 .
- The luxury auto maker last year sold
- 3 1,214 cars in the U.S.
- BELL INDUSTRIES Inc. increased its
- 4 quarterly to 10 cents from seven cents a share .
- 5 The new rate will be payable Feb. 15 .

highlight sentences with disagreement

Annotation



User: admin



User: anno1



Built-in layers vs. custom layers

- **WebAnno offers various built-in annotation layers**
 - User can immediately start annotating
 - Only linguistic layers
 - Layer semantics are known
- **Custom layers allow WebAnno to be adapted to unforeseen tasks**
 - Adapt to non-linguistic annotation tasks
 - Adapt to unforeseen linguistic annotation tasks
 - Layer semantics are unknown
- **Import/export of annotated data**
 - Layers with known semantics convert from/to many formats (TCF, CoNLL, ...)
 - Layers with unknown semantics convert from/to generic formats (XMI, ...)



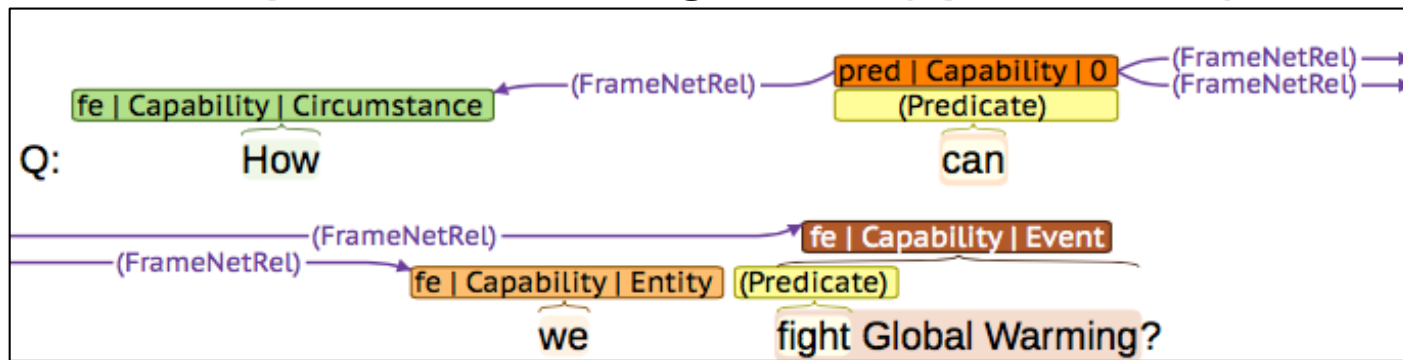
Layer types



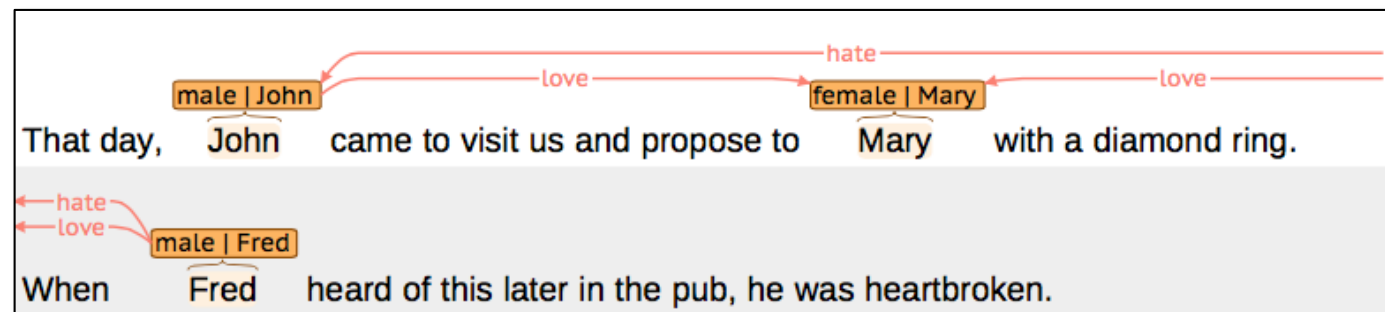
- **Existing built-in layers were generalized into three layer types**
- **Span layer** – POS, lemma, named entity, ...
- **Relation layer** – Syntactic dependencies, ...
 - Attaches to span annotations
 - Directed, reversible arcs
- **Chain layer** – Co-reference chains, ...
 - Undirected arcs
- **Layers can be further customized using “behaviours”**
 - Character-based or token-based
 - Single/multiple token
 - Crossing of sentence boundaries
 - Stacking

Custom layer examples

Semantic predicates and arguments (span/relation)

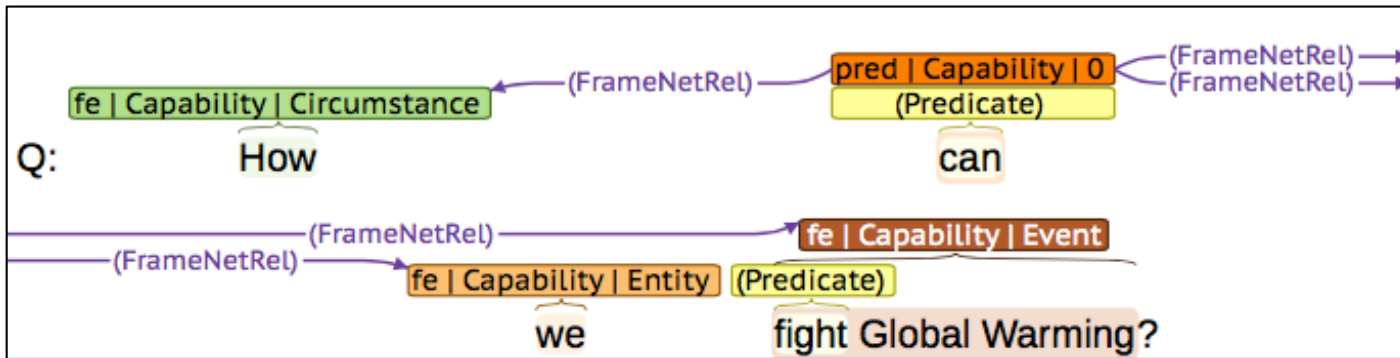


Person (span) / Relationship (relation)



Custom layer examples

Semantic predicates and arguments (span/relation)



Edit Span Annotation

Selected text: **Mary**

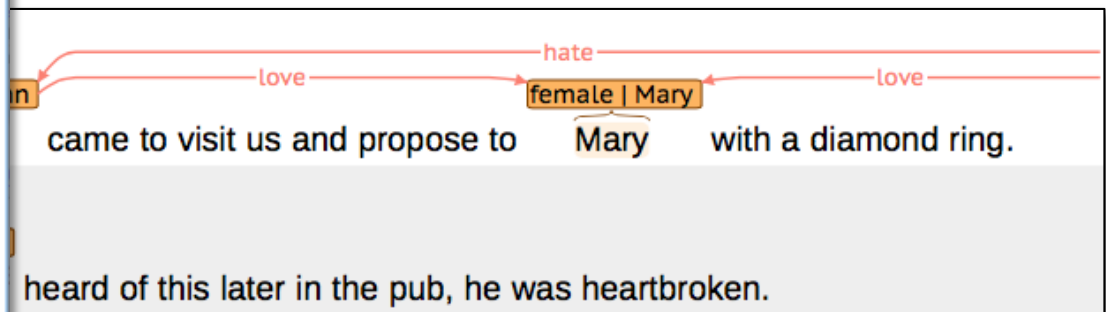
Layer:

Features

gender (Gender):

name:

Person (span) / Relationship (relation)



Custom layer configuration



Details Users Documents **Layers** Tagsets Guidelines Export/Import

Layers

- Coreference
- Dependency
- Lemma
- Named Entity
- POS
- Person
- Relationship

Layers

Create layer

Import layer

Files: Choose Files no files selected

Import layer

Properties ?

Layer name: Person

Description:

Enabled:

Technical Properties ?

Type: span

Attach to layer: -NONE-

Behaviors ?

Lock to token offsets:

Allow stacking:

Allow crossing sentence boundary:

Allow multiple tokens:

Save layer Export layer

Feature overview

gender : [String]
name : [String]

Features

New feature

Feature details ?

Type: uima.cas.String

Feature name: gender

Description:

Enabled:

Show:

Tagset: Gender

Save feature

Control behavior

Controlled vocabulary

Integrated machine-learning



- Annotating data from scratch is more work than correcting
- WebAnno learns from pre-annotated data and makes suggestions
- Accept suggestions with a single click
- Correct suggestions to improve training data



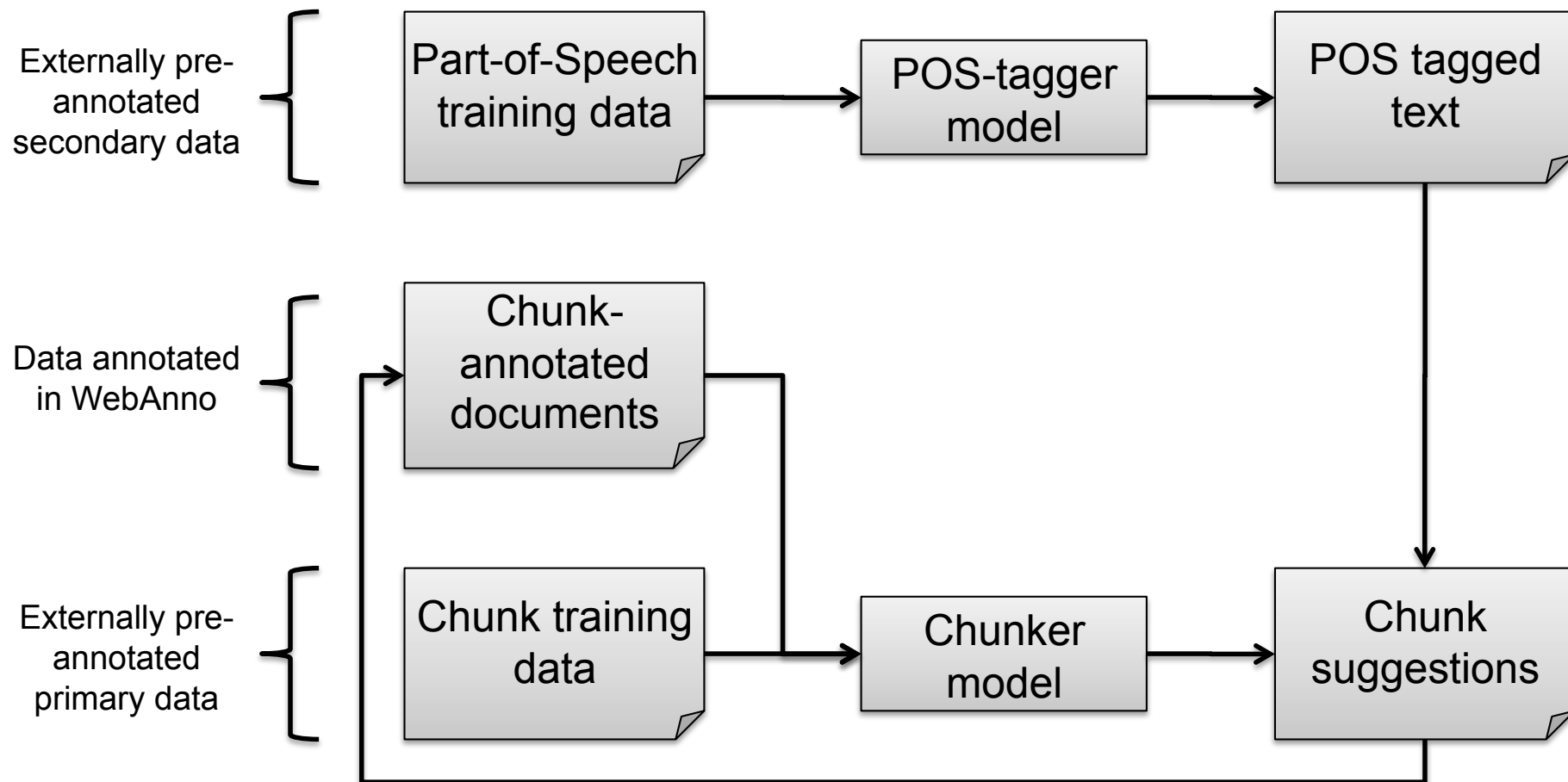
Annotation

1 A blue moon is an extra full moon that appears in a subdivision of a year, either the third of four full moons in a season or, recently, a second full moon in a month of the common calendar.

User: Suggestion

1 A blue moon is an extra full moon that appears in a subdivision of a year, either the third of four full moons in a season or, recently, a second full moon in a month of the common calendar.

Example: Chunking



Automation configuration



Details Users Documents Layers Tagsets Guidelines Export/Import Automation

Select automation layer
[Chunk] value

Start automation

Target layer TAB-SEP target Other layers TAB-SEP feature

Import new documents
Format: TAB-SEP
Files: Choose Files no files selected
Import

Documents
test_chunk.txt
train_chunk.txt
Delete

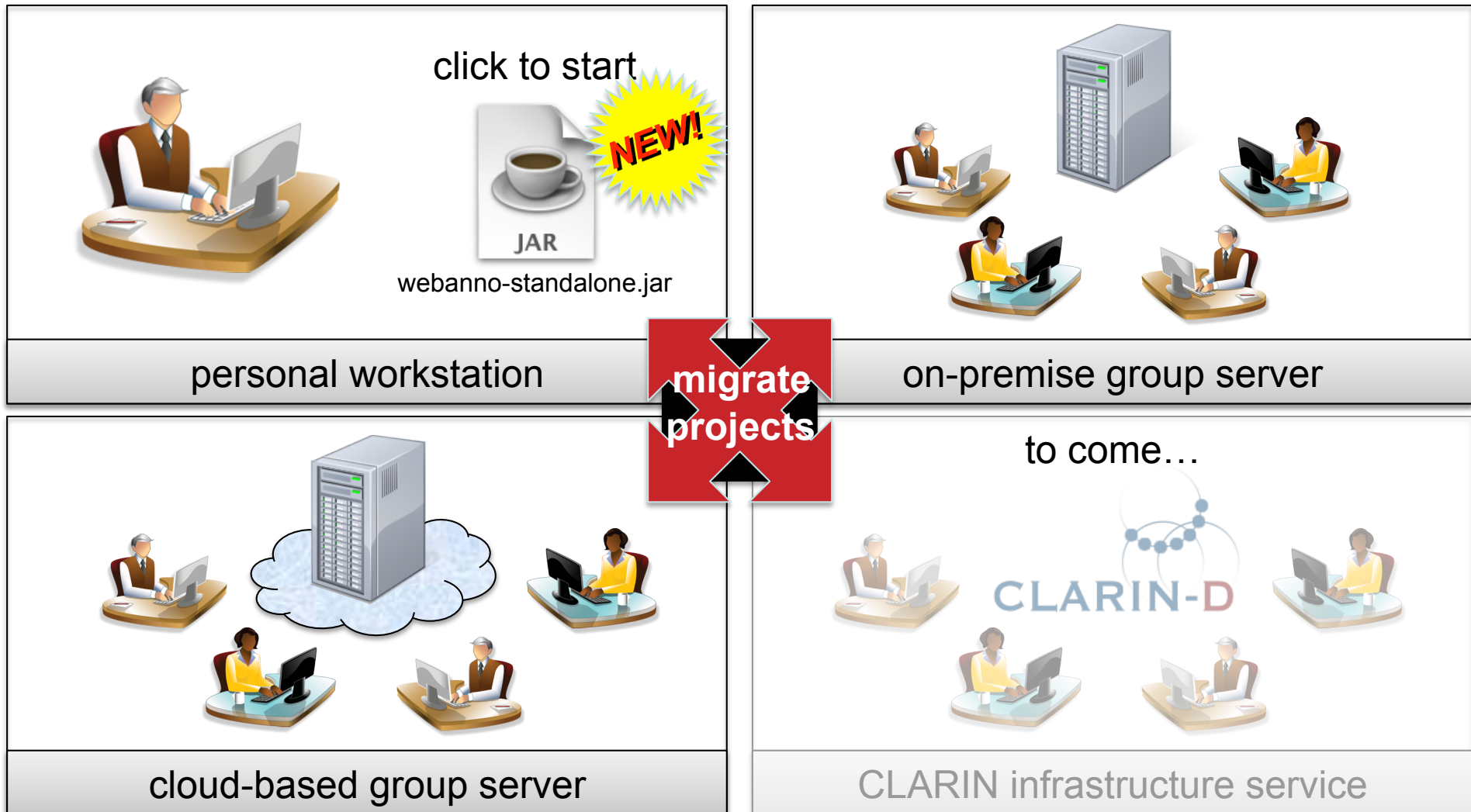
Secondary training data

Primary training data

Training data example

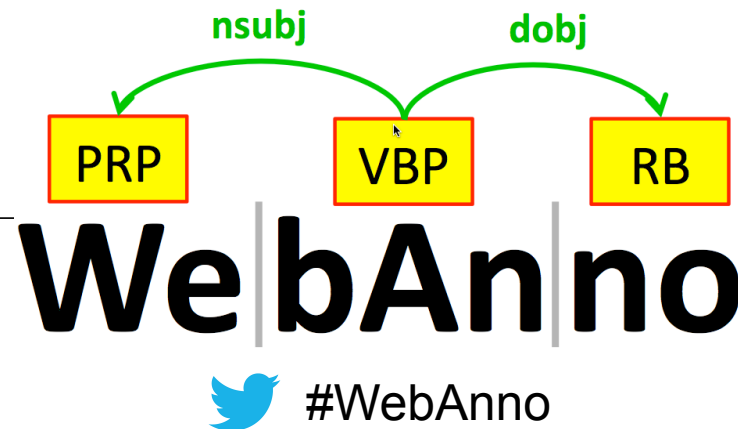
```
test_chun... — Edited
Rockwell B-NP
International I-NP
Corp. I-NP
's B-NP
Tulsa I-NP
unit I-NP
said B-VP
it B-NP
signed B-VP
a B-NP
tentative I-NP
agreement I-NP
extending B-VP
its B-NP
```

Deploy WebAnno as you need it



Where we want to go from here...

- **Extend the scope of WebAnno**
 - Support for slot-based annotation layers (semantic annotations)
 - Tagset constraints
 - Support for more built-in linguistic layers
- **Improve continuously based on user feedback**
 - More efficient annotation interface
 - Support for additional corpus formats
 - ... your feedback?
- **Deploy as a CLARIN infrastructure service**
 - CLARIN AAI support
 - Reduce administrative overhead for operators
 - Self-service for project managers



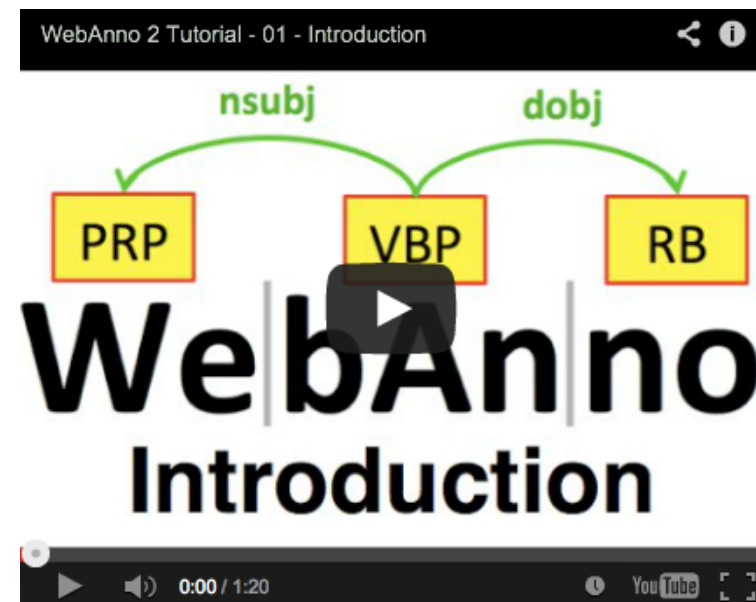
Visit me in the
demo session!

<http://webanno.googlecode.com>

WebAnno Community Survey

This survey is designed to help us understand how people use WebAnno or plan to use WebAnno. So if you are not already a user of WebAnno, we still appreciate your feedback. In this case, please answer the questions such as they reflect your intended use. If you find any of the questions not to be applicable, please simply leave it empty.

Thank your for participating in our survey!



WebAnno 2 Tutorial - 01 - Introduction

PRP VBP RB

nsubj dobj

WebAnno

Introduction

0:00 / 1:20

YouTube