#### CIVET

# Contentious Incident Variable Entry Template: Where we are, what should we do next?

Philip A. Schrodt

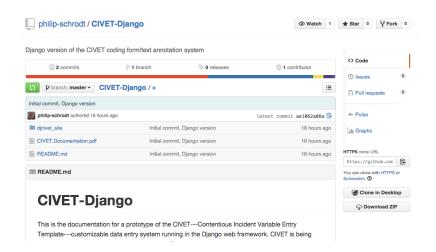
Parus Analytics Charlottesville, Virginia schrodt735@gmail.com

Presentation at Odum Institute, University of North Carolina at Chapel Hill 13 July 2015

#### Developments since March

- Switched from Flask to Django framework
  - ▶ Built-in supervisor/user authentication
  - Django interfaces with a mySQL database
  - But consequently requires more resources and cloud deployment is more difficult
- ▶ Defined a full document format in YAML
- ▶ Used "ckeditor" to create a annotation/editing system
- ► Implemented coder/extraction system to work with the annotation

# Accessing the code https://github.com/philip-schrodt/CIVET-Django



#### Installation on Macintosh

- 1. In the Terminal, run sudo pip install Django
- 2. Download the CIVET system from https://github.com/philip-schrodt/CIVET-Django, unzip the folder and put it wherever you would like
- 3. In the Terminal, change the directory so that you are in the folder *Django\_CIVET/djcivet\_site*
- 4. In the Terminal, enter python manage.py runserver
- In a browser, enter the URL http://127.0.0.1:8000/djciv\_data/

#### At which point you should see...

#### CIVET

#### Contentious Incident Variable Entry Template



This is the home page for a prototype of the CIVET —Contentious Incident Variable Entry Template—data entry system. CIVET is being developed by the NSF-sponsored project titled "A Method for Leveraging Public Information Sources for Social Science Research" which is creating tools to improve the efficiency of data generation in the social sciences, with an initial focus on coding event data in the domain of contentious political.

The system you are looking at is a rather ugly prototype and provides a testbed of testing the functionality of some of various elements we expect to have in the final system. It will also be evolving as we add additional elements.

The prototype is currently deployed as a "Django" (https://www.djangoproject.com/atart/overview/) application; it should be possible to get; this working by installing Django on a local machine and copying the directory djelove\_site. We currently expect the operational version will be deployed on a UNC server though the code will be available if you wish to create your own local or cloud versions.

#### Click one of the following links to use the system

Basic coding template

Template with annotation and extraction

Download documentation (PDF)

#### Acknowledgements

The development of CIVET is funded by the U.S. National Science Foundation Office of Multidisciplinary Activities in the Directorate for Social,

Behavioral & Economic Sciences, Award 1338470 and the Odum Institute at the University of North Carolina at Chapel Hill with additional

assistance from Pariss Analytics.

Last update: 10 July 2015

# CIVET component "layers"

- ► L0: log-in/authenication Status: not implemented but will use the existing Django facilities
- ▶ L1: Translation of raw texts into YAML format Status: prototypes for Factiva
- ► L2: Reading/writing YAML files Status: fully implemented except for audit trail
- ► L3: Sorting texts between "collections" Status: prototyped in Flask
- ► L4: Annotation/editing Status: fully implemented
- ► L5: Coding/extraction Status: implemented except for linkage to new categories

# YAML Components

- ► Collection: Sets of related texts Meta-data: date, comments
- ➤ Texts: Individual texts in original and annotated form Meta-data: source, publisher, license, author, geographical location, comments
- ► Cases: variables coded from this collection Meta-data: coder, date coded, comments

#### YAML Example

```
collid: TestTexts_001
colldate: 2015-06-08
colledit: 2015-06-08
collcmt: Text file source: Apr15.0TH.stories.txt
```

#### texts:

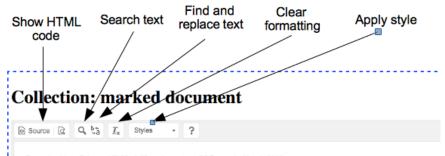
- textid: TextTexts, 002\_001
textpublisher: BBC Monitoring Africa
textpublisher: BBC Monitoring Africa
textpublisher: BBCAP00020150401e04100105
textgeogloc:
textlang: English
textling: English
textlings: English
textlicense: (c) 2015 The British Broadcasting Corporation. All Rights Reserved. No material may be reproduced except
with the express permission of The British Broadcasting Corporation.
textlede: Several people killed in clashes along Chadian-Libyan border
textent:
textoriginal: |
Text of report by Chadian privately-owned. pro-government website Alwinda on 30 March

Several people have been killed and others injured in violent clashes in Gotrone along the Chad-Libyan border. According to the latest information collected by AlwindaInfo, since yesterday [29 March], a heavily armed group from Toubous is said to have been in hot pursuit of assailants who attacked one of their vehicles, leading to the death of two of their people.

#### textmkup: |

div class="textblock" data-textide" TestTexts\_002\_001"xdiv class="textLede" style="color:green; font-weight: bold;">
Several people killed in clashes along Chadian-Libyan border</div<div class="textLeder" style="class:nament; color:blue">Chadian</div</div<div class="textLeder" style="class:nament; color:blue">Chadian</dyspan> privately-owned, pro-government website
<span style="class:nament; color:blue">Chadian
/span style="class:nament; color:blue">Namich Several
/span style="class:nament; color:blue">Namich Several
/span style="class:nament; color:blue">Namich Several
/span style="class:nament; color:red" title="whacked">xilled</pan> and others
<span style="class:nament; color:blue">Sourone
/span style="class:nament; color:blue">Sourone
/span style="class:nament; color:blue">Sourone
/span style="class:nament; color:blue">Sourone
/span style="class:nament; color:blue">Namich Several

## ckedit: Annotation and Editing



Rwandan Hutu fighters kill 22 civilians in eastern DRCongo in March 2015

Text of report by DRCongo's UN-sponsored Radio Okapi on 31 March [Presenter] At least three people have been reported to h March in three different villages of likobo groupment by armed men, likobo is situated to the northeast of Walikale territory near the Local rights organization activists said the recent killings increased the number of those who have so far been killed or kidnapped in the rights activities accurate an acquaint figure at the SDID (Department for the Libertian of Department figure at the company).

#### Coding from Annotated Text

# Document: Rwandan Hutu fighters kill 22 civilians in eastern DRCongo in March 2015 Text of report by DRCongo's UN-sponsored Radio Okapi on March March 2015 Text of report by DRCongo's UN-sponsored Radio Okapi on March Presenter] At least three people have been reported to have been killed sunday March in three different villages of Itolog outpument by armed men. Robo is situated to the northeast of Walikale territory near the border with Lubero territory. Lead rights organization activities said victims were respectively from Kalehe, Kamme and Rusamambo villages. Here is a report filed by Bernadin Nyangi. (Reporter) Local sources said the recent killings increased the number of those who have so far been killed or kidnapped in the entity this month to March Text of the Charleston of Rwandio I fighters under the command of a certain Mutoka, as perpentages in the killings have have consumed of a certain Mutoka, as perpentages in the killings have here as well as the pass of the sense of the charleston of Rwandio I fighters under the command of a certain

the same FDLR faction of having raped at least women this month alone as dother individuals were victims of torture and other sorts of

human rights abuses.

#### Template form for CIV-yaml demo files

	Type of incident: • Demonstration One-sided Violent Armed Clash
	Nature of incident Minor mischlef
	If "Other", provide details in the report section
	Did incident involve local authorities?
s	Location Enter names here
	Maximal injuries Enter names here
	Brief description of incident
	Enter brief description here
	da da

Options after saving: Return to this collection Select new collection Download data

## Extracting Specific Types of Information from Annotated Text

## Rwandan Hutu fighters kill 22 civilians in eastern DRCongo in March 2015

Text of report by DRCongo's UN-sponsored Radio Okapi on March [Presenter] At least three people have been reported to have been killed on Sunday March in three different villages of [kobe] groupment by armed men. Ikobo is situated to the northeast of Walikale territory near the border with Lubero territory. Local rights organization activists said victims were respectively from Kalehe, Kanune and Rusamambo villages. Here is a report filed by Bernadin Nyangi. [Reporter] Local sources said the recent killings increased the number of those who have so far been killed or kidnapped in the entity this month to [March 1]. The rights activists pointed an accusing finger at the FDLR [Democratic forces for the Liberation of Rwanda] fighters under the command of a certain Mutoka, as perpetrators in the killings and kidnappings. They accused the same FDLR faction of having raped at least [March 2] women this month alone as [March 2] of the properties of the same for the rindividuals were victims of torture and other sorts of human rights abuses.

# Template form for CIV-yaml demo files

Type of incident: • Demonstration One-sided Violent Armed Clas
Nature of incident Minor mischief
If "Other", provide details in the report section
Did incident involve local authorities?
Location Enter names here
Maximal injuries Enter names here
Brief description of incident
Enter brief description here

## Remaining steps to reach beta 1.0

- ► Authentication Status: not written but Django has this built in
- ► Read/write sets of collections as zipped files Status: code written but not integrated
- ► Audit trail
  Status: not implemented but everything has been written
  with this in mind
- Specifying customized sets of annotation terms
   Status: prototyped but not integrated
- Sorter
   Status: very ugly Flask prototype; probably needs to be re-written
- Documentation and training videos Status: work in progress

# Key open question: how will this be deployed?

- ▶ Individual system: fully operational on Mac OS-X; still needs testing on Linux and Windows but this should mostly be an issue of getting Django installed
- ▶ Cloud: Deploying on Google App Engine is proving to not be straightforward but other systems might be
- ▶ Server at Odum: do we need this?
- ▶ Multiple-user/coding-farm server at PI institution: Are there general solutions here?

## Additional design issues

- ▶ Persistent vs. transient data: should the data remain on a server or always use upload/download?
- ► Turn-key vs. model code: Are we better off with a more limited but well-documented system that can be used "off-the-shelf" or a more complex system that will usually require some additional customization?
- ▶ Additional features vs. additional documentation vs. making it look pretty
- ► Anyone ready to be a [supported] guinea pig for this?

  "The early bird gets the worm but the second mouse gets the cheese"

#### General categories of additional features - 1

For additional details, see 12 July 2015 memo "Prioritizing features for Civet (Contentious Incident Variable Entry Template)"

- ▶ Document and work-flow management utilities
  - ▶ Formatting source texts into YAML collection format
  - ▶ Automatic sorting and classification
  - ▶ Post-processing utilities, e.g. multiple output formats, reliability and consistency checks
  - Allocating texts to coders
- Look and feel
  - ▶ Make it pretty
  - Maintain the basic system in Flask?
  - ► Hide/show fields
  - ▶ Conditional fields in forms

## General categories of additional features - 2

- ▶ Automated annotation
  - ▶ Dates, which are complicated
  - ► Regular expressions
  - ▶ Geolocation
  - ▶ Numerical equivalents to words: "ten", "two hundred", "many", "dozens"
- Coding form
  - ▶ Additional HTML5 fields for numbers and dates
  - ▶ Local and remote name and code standardization
  - ▶ Templates which automatically fill in fields
  - ▶ Pattern-based and/or dynamic "best-guess" completion
  - Consistency checking

#### Thank you

```
Email: schrodt735@gmail.com
```

Slides:

 $\verb|http://eventdata.parusanalytics.com/presentations.html|$ 

Software:

 $\verb|https://github.com/philip-schrodt/CIVET-Django|\\$