API-Powered Dictionaries
For Digitally Under-Represented Languages

Sandro Cirulli
Oxford University Press (OUP)

FOSDEM
31 January 2016
Table of contents

1. Introduction

2. API

3. Demo

4. How to contribute

5. Summary
Less than 5% of the current world languages are in use online

Oxford University Press (OUP)

- Oxford University Press (OUP) is a world-renowned dictionary publisher

- OUP launched the Oxford Global Languages (OGL) initiative to digitize under-represented languages

- In September 2015 OUP launched two African languages websites for isiZulu and Northern Sotho
Oxford Global Languages (OGL)

Vision

- Localized websites and digital access for multiple languages
- Language communities contributing content (crowdsourcing)
- Supporting digitally under-represented languages
- Flexible data serving multiple needs
isiZulu and Northern Sotho websites powered by APIs

Translation of *freedom* in Northern Sotho

**freedom**

**NOUN**

**tokologo**

**Examples**

He was the first South African person to fight for freedom

_E be e le yena motho wa mathomo wa Afrika-Borwa wa go lwela tokologo_*

Newly added words

- **snack**
  - *adodomatele*

- **yesterday**
  - *adodomatele*

- **tlhohlo**
  - *mmunhuru494*

- **committed**
  - *mmunhuru494*

How do you say **literacy** in Northern Sotho?

How do you say **snack** in Northern Sotho?
What is an API?

- API stands for Application Programming Interface
- A (Web) API is a set of rules for exchanging information with a website
- An API is a machine-to-machine interface for receiving and sending data via HTTP requests
RESTful API for OGL websites

1. User:
   - fill web form
   - show confirmation message

2. Website:
   - request entry
   - send POST request
   - send HTTP code
   - show confirmation message

3. API:
   - send SPARQL query
   - return HTTP code
   - send SPARQL query
   - send GET request
   - serialize RDF into JSON-LD
   - return SPARQL result

4. Triple Store:
   - run SPARQL Update query
   - run SPARQL query

Diagram illustrates the sequence of events involving user interaction with a website, API requests, and data storage in a triple store.
Benefits of API

- **Reusability**: data for other languages can reuse the same API thus reducing costs

- **Flexibility**: data can be shipped in multiple formats (XML, JSON, JSON-LD, RDF, etc.)

- **Integration and Automation**: external systems, applications, and developers can easily integrate and consume data
How to contribute

▶ Contribute new content to isiZulu, Northern Sotho, Urdu, and Malay dictionaries - more languages in the next months!

▶ Try out our API

▶ Experiment with our SPARQL endpoint at https://github.com/OUP-DTG/sparql_endpoint

▶ We are recruiting!
Summary

- OUP launched the Oxford Global Languages (OGL) initiative to **digitize under-represented languages**

- isiZulu and Northern Sotho dictionary websites are **powered by APIs**

- **Contribute** to the OGL initiative and try out our API and SPARQL endpoint
Thank you for your attention!

Contact:
www.sandrocirulli.net/contact
sandro.cirulli@oup.com

Slides:
www.sandrocirulli.net/fosdem2016

Links:
OGL programme: www.oxforddictionaries.com/ogl
isiZulu website: zu.oxforddictionaries.com
Northern Sotho website: nso.oxforddictionaries.com
SPARQL endpoint: github.com/OUP-DTG/sparql_endpoint