A WIKI APPROACH TO COLLECTING, CURATING AND MANAGING CITIZEN SCIENCE DATA

Daniel Mietchen

Data Science Institute, University of Virginia, `daniel.mietchen@virginia.edu`, @EvoMRI

Wikimedia-based communities

Over 100 organizations have been founded around the globe to support Wikimedia-based collaboration in specific local, regional, thematic or technical contexts.

Participatory science

The key aspect of participatory science - often referred to as citizen science, though that term is problematic - is the participation of volunteers in research projects. In principle, this can happen at any stage of the research process.
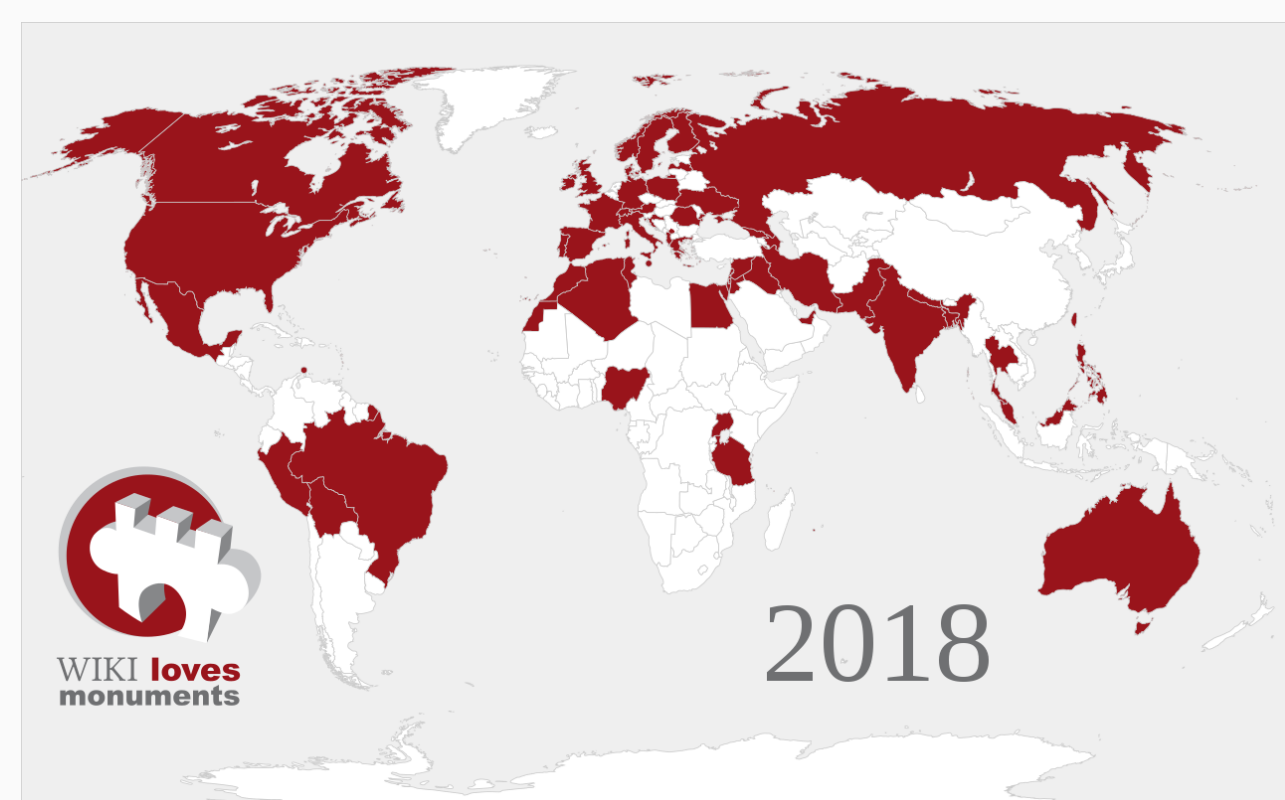
Wikimedia projects

Wikimedia projects include Wikipedia and its sister projects, which are basically organized around information channels, i.e. Wikipedia as the encyclopedia, Wikibooks for books, Wikimedia Commons for media, Wikidata for data, Wikisource for source materials, Wikivoyage for travel guides and so on. All of these projects use the MediaWiki software, most of them exist in multiple language versions, and they collaborate in various ways across projects, languages, disciplines, jurisdictions etc.

Wikimedia projects as citizen science platforms

Besides *writing about research*, Wikimedia projects can in principle be used within any open research project but use in a citizen science project manner is most common, e.g. for the collaborative identification of biological or mineral species, historic personalities or buildings as well as for transcribing or translating documents or for geolocating historic maps.

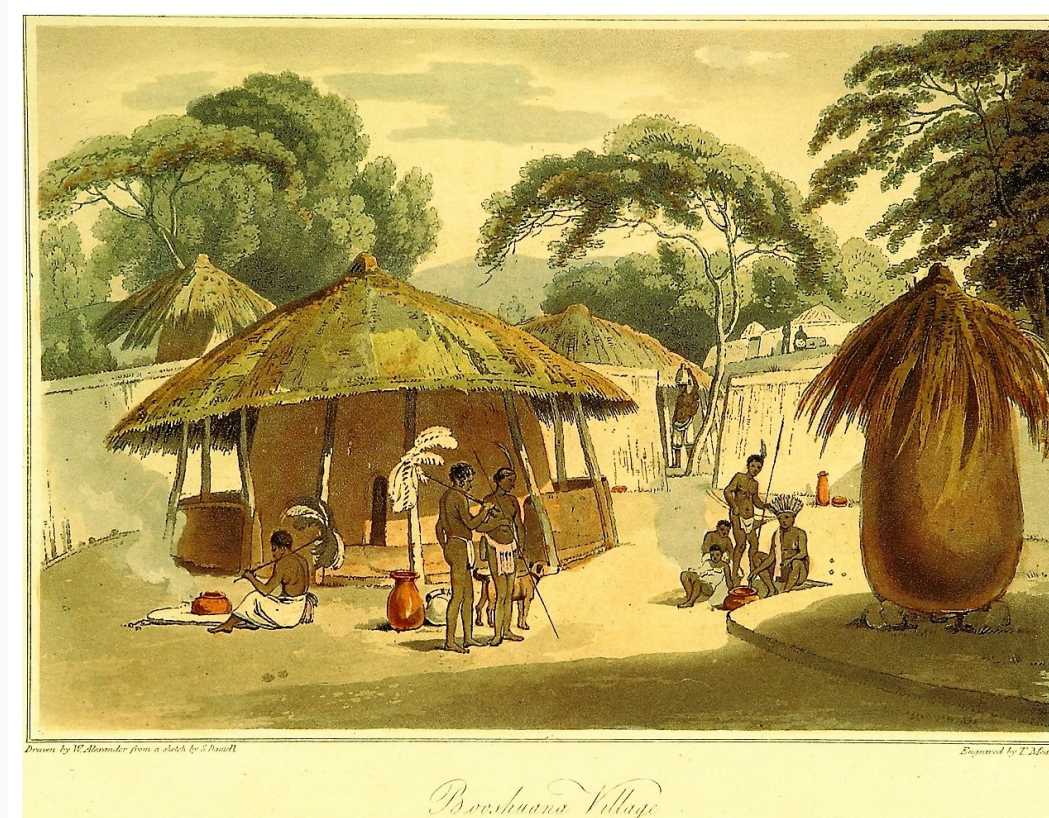## Examples of participatory science on Wikimedia platforms

Wiki Loves Monuments, the largest photography contest in the World, is aimed at documenting cultural heritage

Collecting existing documentation of lost cultural heritage, here after the fire in the Brazil National Museum

Documenting history, e.g. this colonial depiction of a 'Booshuana' village in 1806

This species of *Megistocera* was first discovered on Wikimedia Commons and remains undescribed.

Crowd-designed open infrastructure to bring Wikimedia content to places without Internet access

Wikidata as a participatory science project in digital curation

Wikidata (**https://www.wikidata.org**) is a multilingual collaborative platform that democratizes data curation similar to how Wikipedia democratized the curation of encyclopedic information. It is tightly integrated with all language versions of Wikipedia and its sister sites, and it collects, reuses and provides structured public domain data across all areas of knowledge from all around the world. Wikidata meets the requirements of the FAIR principles to make data findable, accessible, interoperable and reusable, and it allows people to collaborate who do not share a common language. With about 20,000 volunteer contributors each month that collaborate openly, Wikidata blends open science and citizen science approaches.
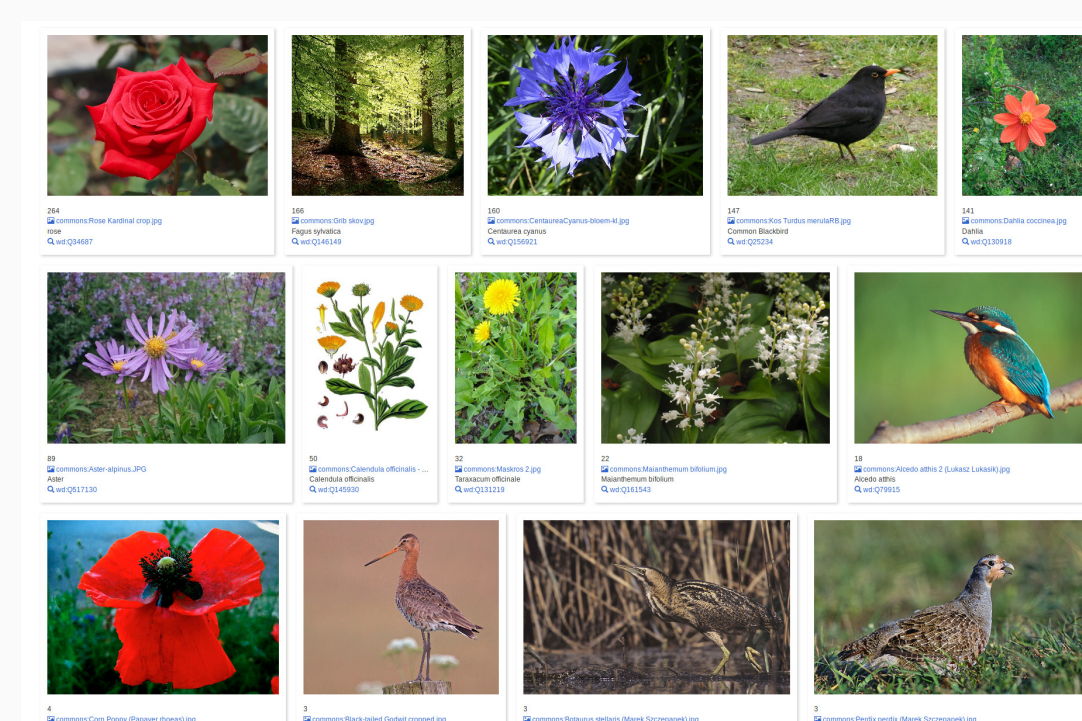
The human contributors are aided by hundreds of automated or semi-automated tools that perform repetitive tasks at scale, based on community-agreed standards. Together, they have aggregated over 6 billion RDF triples on the platform that can be queried via a dedicated SPARQL endpoint and other means, which aids in quality control of the database content and workflows, and facilitates knowledge discovery within the corpus.

Thanks to a combination of extensive examples, help pages, tutorials, user interface design and other mechanisms, this endpoint is gentle to users across various skill levels for the SPARQL query language. This way, Wikidata also democratizes access to and participation in the Semantic Web.
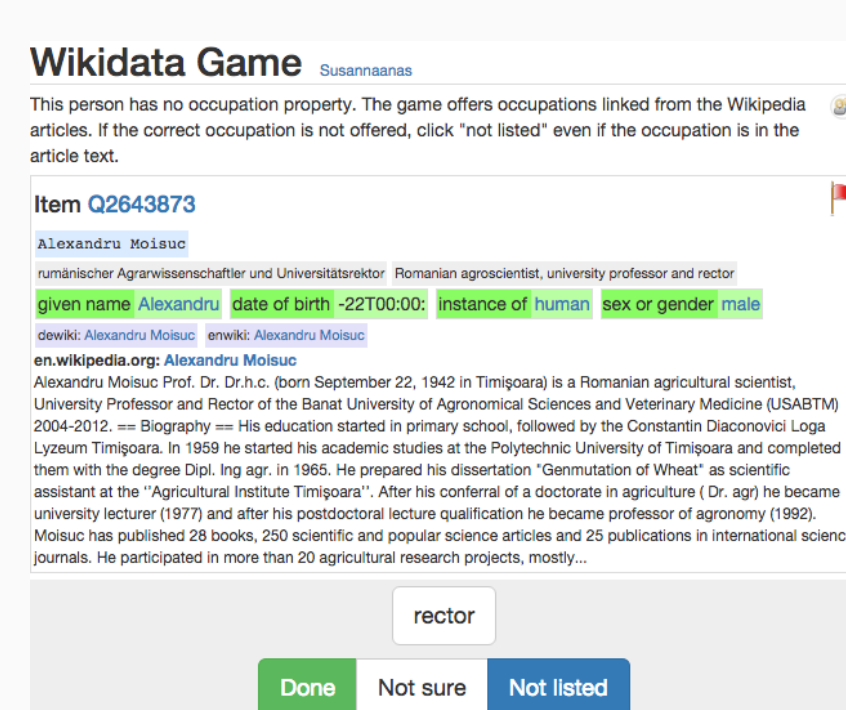
Wikibase as a generic crowdsourcing platform

The software underlying Wikidata is Wikibase (**http://wikiba.se/**). It is open source and openly licensed, which allows anyone to run semantic databases that are interoperable with Wikidata and other Wikibase instances. By default, Wikibase instances come with a SPARQL endpoint of their own that is modelled after the Wikidata one.
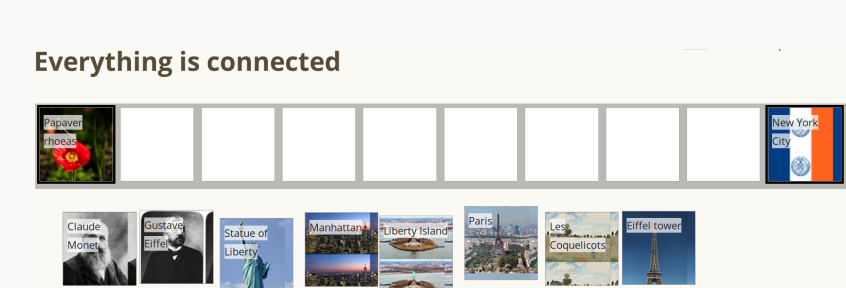
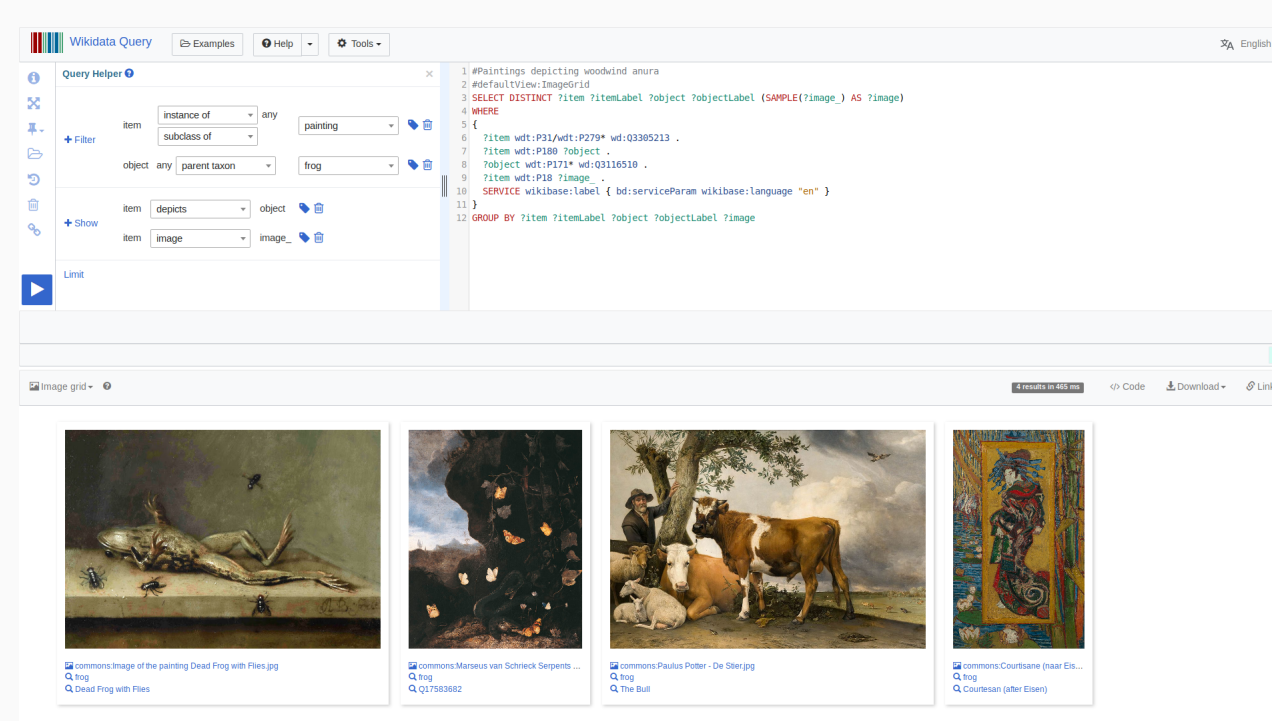## Examples of participatory science on Wikidata or Wikibase

Mapping streets in the Netherlands in terms of the species they are named after
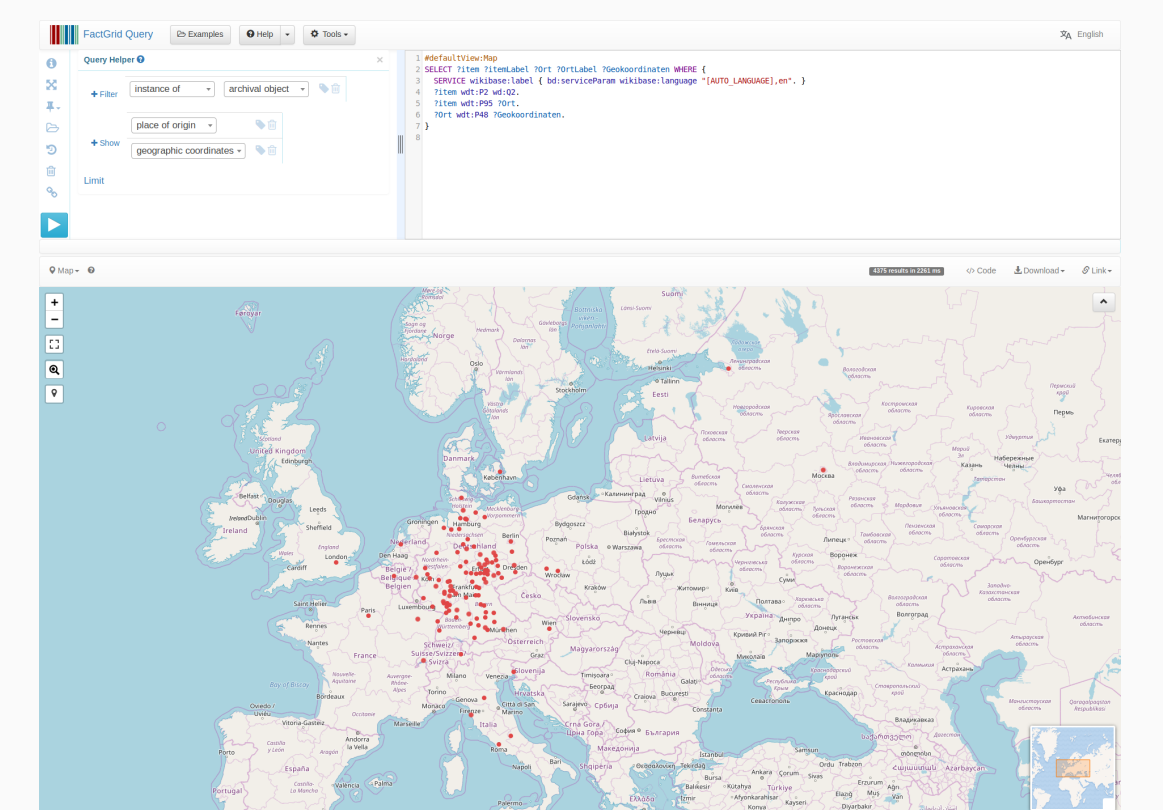

Wikidata-based game

Wikidata-based puzzle

Results of a Wikidata SPARQL query for 'Paintings depicting frogs', based on collaborative annotation of paintings

Community page for help with Wikidata SPARQL queries

Results of a query to the SPARQL endpoint of FactGrid, a Wikibase instance, depicting locations from which Illuminati members have sent letters