Enriching Cultural Heritage Collections with Crowd Generated Knowledge

Jasper Oosterman[1], Alessandro Bozzon, Geert-Jan Houben
Archana Nottamkandath, Chris Dijkshoorn, Lora Aroyo
Mieke H. R. Leyssen, Myriam C. Traub
[1] j.e.g.oosterman@tudelft.nl

Delft University of Technology
VU University Amsterdam
Centrum Wiskunde & Informatica

Challenge — Cultural Heritage Collections

Enrich data collections by tapping into the **interest** and **expertise** of **crowds** to create knowledge; **Crowd Generated Knowledge**.

- ✓ Data-intensive
  - Rijksmuseum has 1M art pieces requiring annotation

- ✓ Knowledge-intensive
  - Diverse and specific knowledge needed

- ✓ Goals
  - Coverage: Enrich complete (sub)collection
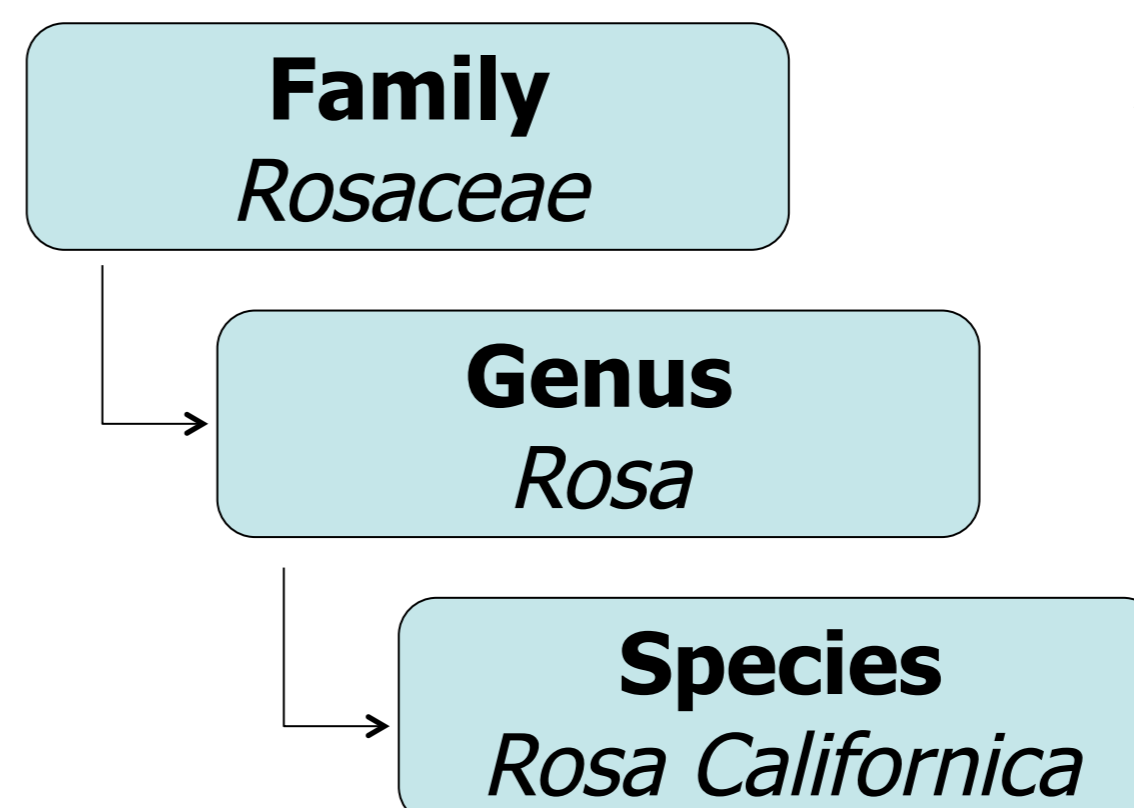 - Quality: High quality annotations

1 — Aspects of Knowledge Intensive Tasks

Identification
- Identify relevant entities
- Prominence and amount of entities
- Artistic interpretation, lack of detail, fantasy

Family
Rosaceae

Genus
Rosa

Species
Rosa Californica

Annotation
- Tag identified entities
- Specificity of tags
- Domain and culture specific knowledge

How can experts from the crowd support the enrichment of Cultural Heritage collections?

Links — Demo — Video

Experiment performed within the SEALINCMedia project. Scan for a demo of Accurator or a video explaining our research together with the Rijksmuseum.
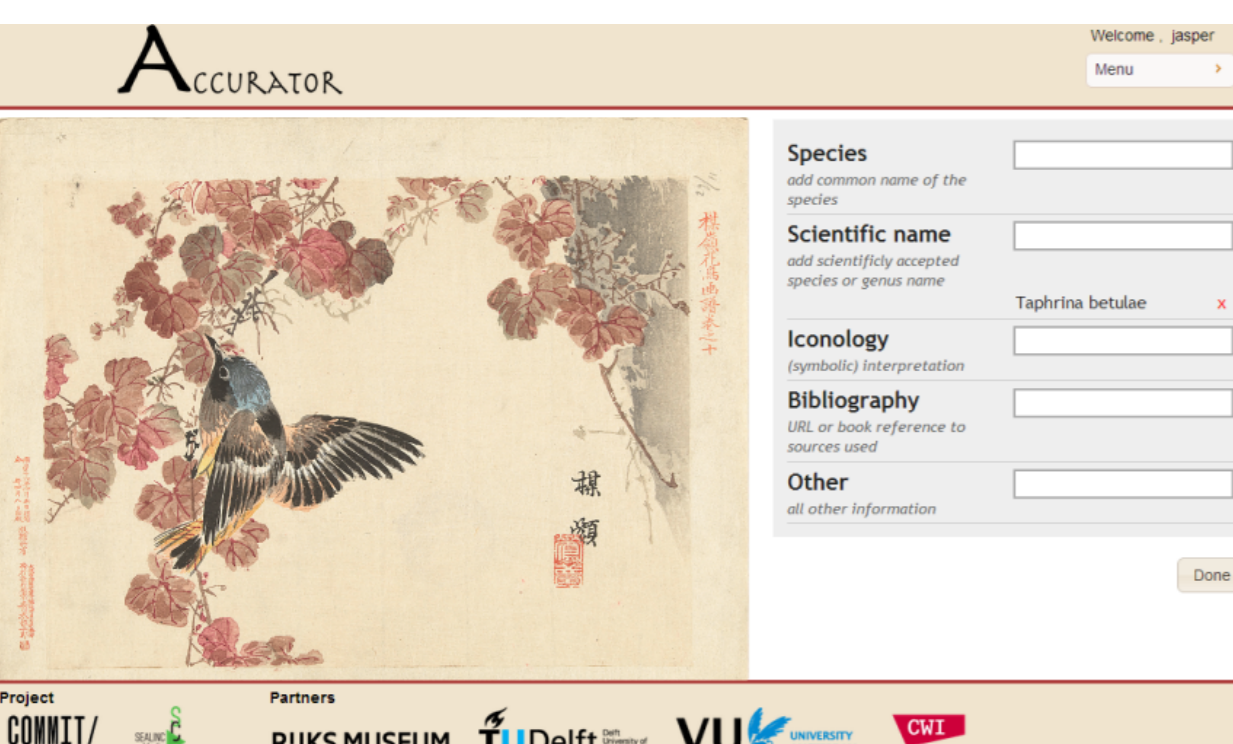
2 — Experiment

Setup
- 86 prints from the Rijksmuseum containing flowers
- Tasks: annotate prints with specific flower names
- Executed by experts and crowd workers via crowdsourcing platforms
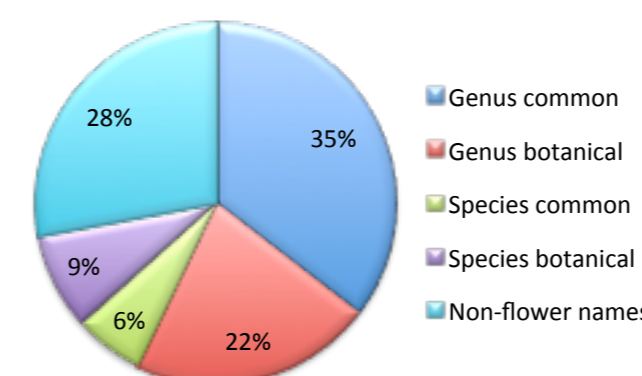
Experimental platform: Accurator

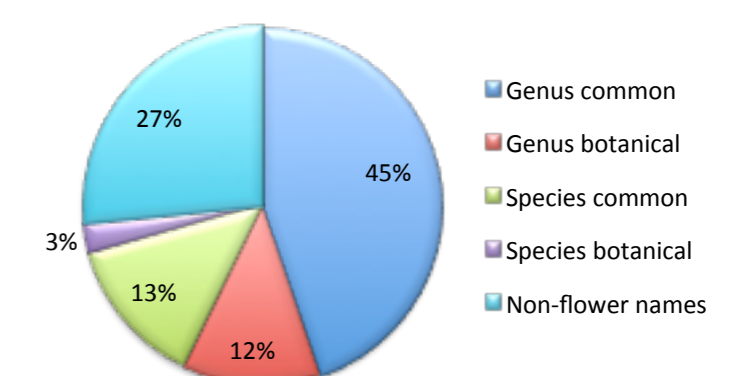A platform to support crowd-enabled, collaborative annotation processes.

3 — Conclusions

Number of	Experts	Crowd workers
Annotators	4	75
Annotation tasks performed	128	982
Tags (total)	161	1077
Tags (flower names)	119	831
Distinct flower names	47	123 (17 overlapping names)

Type of experts annotations

35% Genus common, 22% Genus botanical, 6% Species common, 9% Species botanical, 28% Non-flower names

Type of crowd workers annotations

45% Genus common, 12% Genus botanical, 13% Species common, 3% Species botanical, 27% Non-flower names

Insights
- Botanical flower knowledge is present in the crowd…
- … but image difficulty affects crowd annotation quality
- Low crowd annotator agreement → worker selection and task orchestration are required