

Crowd Generated Knowledge

Jasper Oosterman



E: j.e.g.oosterman@tudelft.nl
T: +31 (0)15 27 86346
@joosterman

Supervisors
Alessandro Bozzon
Geert-Jan Houben

Web Information Systems
Delft University of Technology
http://wude.nl

Motivation

Systems like Wikipedia or online museum collections are data driven. The data itself (a concept/artwork) does not bring success. Value comes from additional **knowledge** about the data. For example: the name of the depicted bird.



Daurian Redstart

Application owners have limited time and/or expertise and cannot provide required knowledge. Other sources of **knowledge creation** are needed.

To enrich data collections we tap into the interest and expertise of **Crowds** to create knowledge. This is **Crowd Generated Knowledge**.

Our research concern the **scientific understanding** of the process to produce Crowd Generated Knowledge to enrich data collections.

Research Overview

Our main research question is

How can we maximize the creation of Crowd Generated Knowledge?

Producing Crowd Generated Knowledge involves **three** steps:

- **Specify** required **knowledge** and creation **constraints**
- **Identify** and **assemble** crowds for expertise finding
- **Plan** and **execute** crowd activities for knowledge creation

Goal specification

Crowd identification

Activity planning

Goal specification

Goal specification requires **two** steps.

Goal elicitation

Knowledge
Stories about Delft painters

Process
Within 2 months

Crowd
Written in Dutch

Quality
High readability

Goal formulation

```
<?xml version="1.0" encoding="UTF-8"?>  
<goal>  
...  
</goal>
```

Use case

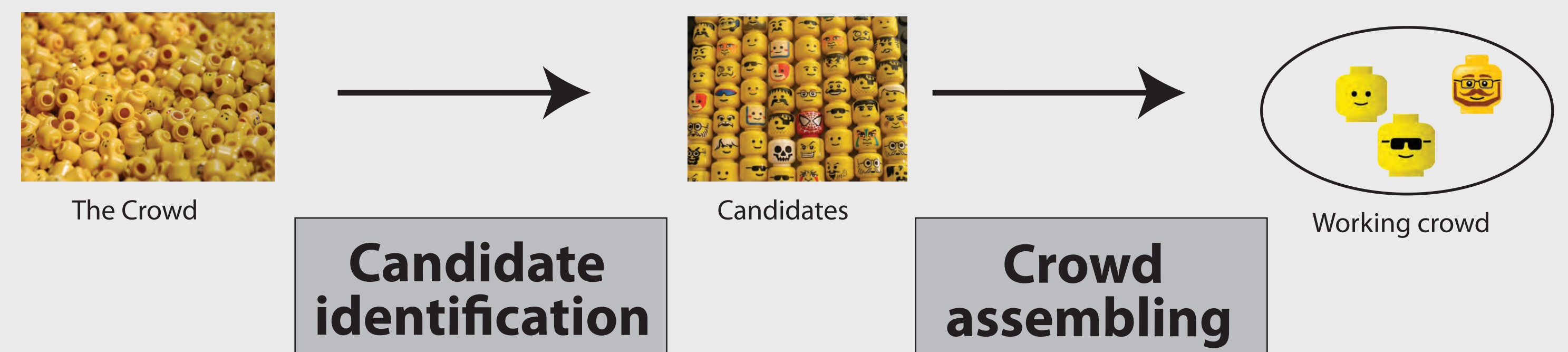
- OpenImages platform contains audiovisual data
- Stakeholder wants improved recommendations
- Required knowledge and constraints (i.e. goals) yet unknown
- Challenges: Elicit and formulate goals

Research

- Identify and formulate goal elicitation process
- Create human- and machine-readable goal formulation language

Crowd identification

Crowd identification requires **two** steps.



Use case

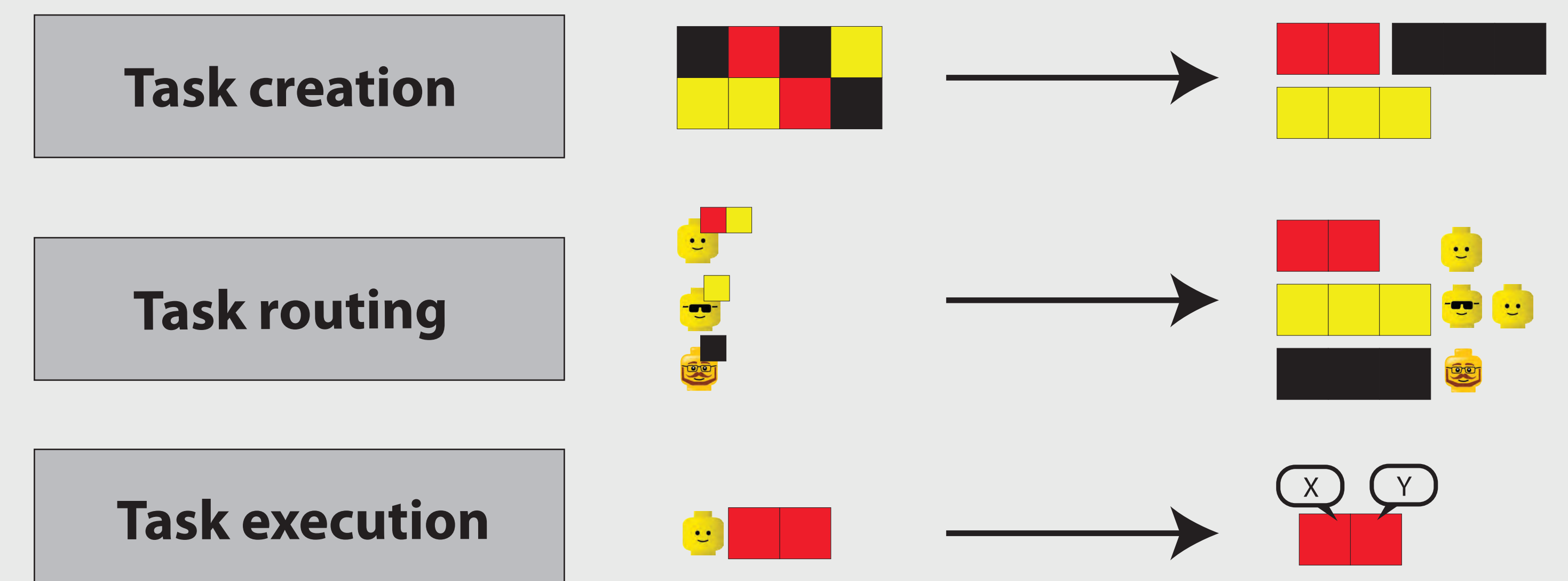
- Stories about Delft-related topics enrich WikiDelft
- Goals: topic coverage and multiplicity of stories
- Challenges: finding the right people to create the story and collaborative story creation

Research

- Modeling knowledge and expertise of crowds
- Create goal driven strategies for crowd assembling

Activity planning

Activity planning requires **three** steps.



Use case.

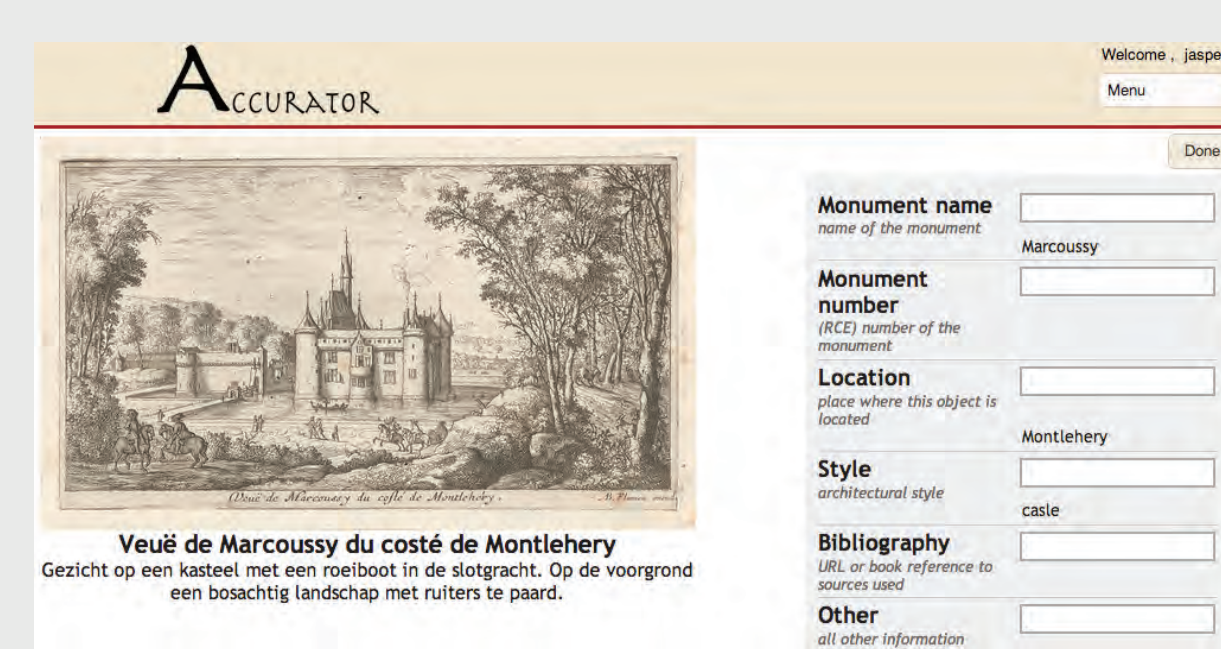
- Rijksmuseum has many prints with unnamed depicted elements
- Specific knowledge about elements is needed
- Goals: coverage of prints and correctness of knowledge
- Challenge: getting specific and detailed knowledge from crowd

Research

- Study the creation of (workflows of) tasks
- Study distribution of tasks to right persons in the working crowd
- Study execution of tasks

Project results

- Reusable open source **software framework** for crowd knowledge generation
- Creation and evaluation of several **candidate identification strategies** on social networks and human computation platforms.
- Creation and evaluation of several **task creation, routing and execution strategies**
- Application to several real world scenarios



Partners



Acknowledgements.

This publication was supported by the Dutch national program COMMIT. This research is affiliated with the Delft Data Science consortium.

COMMIT/