MS10. Training Plan

Version 0.7

Documentation Information

Contract Number	2018-EU-IA-0104
Project Website	https://saintgeorgeonabike.eu/
Contratual Deadline	31/03/2020
Nature	Report
Author	Ariadna Lobo (BSC), Rose Gregorio (BSC)
Contributors	Antoine Isaac (Europeana), Maria Cristina Marinescu (BSC), Joaquim More (BSC), Michelle Lewis (Europeana), Milena Popova (Europeana), Nicholas Jarrett (Europeana), Monica Marrero (Europeana)
Reviewer	Antoine Isaac (Europeana), Maria Cristina Marinescu (BSC), Joaquim More (BSC)
Keywords	Training, workshop, webinar

Change Log

Version	Author	Description Change
V0.1	Ariadna Lobo (BSC)	Initial draft
V0.2	Rose Gregorio (BSC)	Second draft
V0.3	Antoine Isaac (Europeana)	Review
V0.4	Joaquim More (BSC)	Review
V0.5	Maria Cristina Marinescu (BSC)	Review
V0.6	Rose Gregorio (BSC)	Final version
V0.7	Rose Gregorio (BSC)	Modification to training events timeline

Table of Contents

1. Executive Summary

This document is centered around the plan for specialised training and education activities related to improving the quality and quantity of open metadata through AI content analysis technology. In order to train the new generation of researchers and cultural heritage professionals, Saint George on a Bike (SGoaB) will offer a sustained, high-quality training and education programme in the form of different events such as workshops, tutorials and webinars.

The purpose of this Training Plan is to reach a wide community, both academics as well as specialists in the management of cultural heritage data and Public Open Data, with a specific focus on creating a training programme at a European level. In order to achieve this, both [Barcelona Supercomputing Center (BSC)](#) and [Europeana](#) will collaborate to organise the project´s training activities.

2. Introduction

SGoaB intends to devote an important amount of resources and effort in order to train primarily researchers and cultural heritage professionals about the project´s data training methods, implementation, and possible impacts. Moreover, the project will contribute, align, and integrate its training activities with existing European initiatives. This includes exploiting the existing training channels and platforms of other related projects such as [Partnership for Advanced Computing in Europe (PRACE) training centres](#) to participate actively with the suggested training events, and trying to accommodate the content to the course requirements. Another platform where SGoaB training activities will be organised is existing Europeana events.

The training activities will consist of both face-to-face training events as well as webinars, allowing the project to reach a larger audience. They will be communicated to INEA.

3. Target audiences

SGoaB´s training plan is two-tiered and considers target audiences for the research and implementation/applications stages of the project. The primary target audience (for the research/prototyping stage of the project) consists mainly of:

- Academic researchers in data mining, data science and AI
- Cultural heritage professionals working with (meta)data such as metadata experts, data scientists, catalogers, digital library project managers, etc.

The secondary target audience (for the implementation stage of the project) are:

- Cultural heritage professionals such as curators, decision-makers
- Educators

The training and support events will address different member states, with the goal of reaching at least 10 member states every year (e.g. by addressing at least 5 member states per training and supporting session).

Particular attention will be paid to training young people and women researchers working in research and management of cultural heritage data. In order to maintain gender balance, an equal ratio of men and women instructors will be encouraged when organizing the training courses.

3. Training formats

A minimum of six events are foreseen during the project. They will be in the form of face-to-face tutorials or workshops, as well as online training webinars.

In order to define the contents of the courses, SGoaB first identified the training needs of researchers and participants in the project. Partners were contacted and asked to share what kind of topics they are interested in, think may be interesting and useful, and what they could teach. According to their answers, the following topics should be included in the training courses:

- Production of training metadata
- Crowd annotation
- TimeMatrix: A new method for contextualising image description
- Contextualization for creativity
- Tools for data search
- Artificial Intelligence for cultural heritage

Face-to-face training events will be structured in two parts: theory and hands-on sessions, with the latter covering at least 50% of the course. Each training course will limit the maximum number of attendants depending on the contents and structure.

The training events will not only be an opportunity to learn; they are also expected to serve as a platform for participants to share their own experiences and challenges. This guided SGoaB to identify some conferences at which the project could organize training events as "hosted events" (tutorials or workshops). This will optimise logistic efforts across the community, help the project target a wider audience and reach out to participants who can also share their own experience and perspective, thus making the event much more effective for all the participants.

For the online training activities (webinars), SGoaB will seek to attach project activities to Europeana's own webinar effort, whereby webinars will be promoted on the Europeana Pro platform. To optimise visibility in the project's target audiences, SGoaB also intends to collaborate with existing community platforms such as the Association of European Research Libraries (LIBER) and Dublin Core Metadata Initiative (DCMI). The webinars would then be recorded and uploaded on the respective websites and a link to them would be shared on the SGoaB and Europeana Pro websites as well.

The training material will be uploaded on the project website before the start of each course.

4. Training events timeline

Based on the topics suggested by SGoaB partners, several have been chosen and further defined in order to develop useful and informative training events for the target audiences.

The table below shows the tentative schedule for these training events. SGoaB plans to offer two training courses in 2020. Due to the uncertainty surrounding the current COVID-19 outbreak, the first training course will be planned as a webinar. This is planned to take place around the middle of 2020.

Date	Subject	Location
Summer 2020	TimeMatrix for Researchers	Online
Fall/Winter 2020	Workshop on AI for Cultural Heritage Image Analysis @ Europeana Conference 2020 (EuropeanaTech track)	TBC
Winter/Spring 2021	Workshop on Tools for Image Related Data Search @ PATC: Big Data Analytics	Barcelona, Spain
Winter 2021	Workshop on AI for Cultural Heritage Image Analysis @ Futurs Fantastiques (AI4LAM)	TBC
TBC	Fitting Crowd Annotation and AI to Cultural Heritage Image Discovery (Wikimedia-related event)	TBC
TBC	Contextualisation for Creativity	TBC

The previous version of MS10 Training Plan (v0.6) indicated that SGoaB would organise a workshop on AI for Cultural Heritage Image Analysis at Futurs Fantastiques 2020. This part has been modified because of AI4LAM´s recent announcement that the event is postponed to 2021. The SGoaB project is still interested in holding a workshop there and will plan accordingly once the details of Futurs Fantastiques 2021 are announced.

5. Event description

5.1 TimeMatrix for Researchers

Description:
The webinar format course will show the SGoaB results on adapting descriptions of paintings and photos to the time period when they were created. The course will pose and discuss challenges for researchers; hands-on activities could potentially be organized for them to suggest possible approaches.

Expected outcome:
Technical improvement of the time machine effect, where the descriptions of the contents of an image can be transformed by deep learning methods depending on the probability of visually representable entities appearing during different time periods.

Target audience:
Cultural heritage professionals working with (meta)data such as metadata experts, data scientists, catalogers, digital library project managers, etc. Academic researchers in data mining and data science that are interested in art and culture.

5.2 AI for Cultural Heritage Image Analysis

Description:
This face-to-face course/workshop aims to explain and illustrate how AI image analysis technologies such as the ones developed by SGoaB can contribute to better understanding and exploiting the past and cultural heritage. Technical and data challenges, solutions and results will

be shown and discussed. Depending on the venue where we organize this workshop, we could extend it with a debate on how AI may affect matters related to social sciences, education, and citizens' cultural awareness.

Expected outcome:
Awareness of latest developments in AI image analysis, hands-on experience with applying existing image analysis tools.
Better perspective on the role of AI in understanding cultural heritage.

Target audience:
Primary audience: academic researchers in data mining, data science and AI, cultural heritage professionals working with (meta)data (metadata experts, data scientists, catalogers, digital library project managers
Potential secondary audience: cultural heritage professionals such as museum curators, decision-makers

We expect to run two iterations of this event, possibly with varying content and targets, at the coming Europeana 2020 Conference, the annual event of the Europeana Network, and Futurs Fantastiques, the annual event of the AI for Libraries, Archives and Museum (AI4LAM) community.

5.3 Tools for Image Related Data Search

Course description:
This course will demonstrate the potential of SGoaB for text-based image search and metadata search. These tools will help researchers in cultural heritage to quickly find relevant material that may not be findable before semantic enrichment, and therefore create a meaningful storyline. This functionality may also be of help to educators when teaching art history and appreciation courses.

Expected outcome:
To assess the usefulness of the search tools for culture and heritage professionals, understand the shortcomings and opportunities for improvement, and maximize the expected impact with these communities.

Target audience:
Researchers in cultural heritage

5.4 Fitting Crowd Annotation and AI to Cultural Heritage Image Search

Course description:
The course will show how useful crowdsourced (or "nichesourced") annotations by curators, educators, students, etc. can be and whether their annotations fit the data search by both common people and targeted groups (e.g., the blind), especially when they consult cultural heritage. The course will also train participants on best practices for annotations of different types: e.g. labels, descriptions, semantic tags. Finally, we will try to set a seed for incentivising our audiences to develop (or respond to) crowdsourcing campaigns in which end users may enrich the data used for either training or validate metadata enrichments.

Expected outcome:

We expect the outcome to be twofold: first we aim to prove the value of crowd annotation for training AI tools to perform well on cultural data, and discuss the feasibility of suitable annotation efforts with potential annotation campaign creators or contributors. As a side effect, but with possibly big implications, we hope to trigger the creation of semantically annotated corpora for training purposes, especially if we manage to involve communities such as Wikimedia and Wikidata.

Target audience:
Wikimedia/Wikidata community members, curators of online cultural heritage data, educators and students in humanities such as art history, history or sociology.

5.5 Contextualisation for Creativity

Course description:
This course will demonstrate all the tools SGoaB will be developing during the course of the project, and how they may be used, separately or in combination, to spark creativity and design within communities interested in cultural heritage: museums, educators, public entities, creative industries, etc. At this point it is still to be determined which event is the best choice to organize a session; this will also depend on the concrete results and tools we will have by Fall of 2021.

Expected outcome:
To involve a wider audience with the results of our project, evaluate the tools with a creative audience, and learn what these communities find our tools useful for, and what they are missing.

Target audience:
Cultural heritage professionals