COURSE INTRODUCTION

TABLE OF CONTENTS

01

INTRODUCTION
course overview

02

BACKGROUND
what you need to learn by your own, what you learn in this course

03

PROJECT
Topic, groups, documentation

04

EVALUATION
Assignments and final exam

INTRODUCTION

IN NUTSHELL

Overview of methods to manage and query data, e.g. statistical analysis

DATA ANALYSIS

Features and techniques to visualise data for multiple purposes

DATA VISUALISATION

How to present data to support exploration, discovery, and decision making

DIGITAL COMMUNICATION

WHAT YOU'LL LEARN

Manipulate data with python.
Access and query RDF data on the web.

DATA ANALYSIS

Choose appropriate visualisations according to your data.

DATA VISUALISATION

Present your results in an engaging way on the web.

DIGITAL COMMUNICATION

BACKGROUND

WHAT I WISH YOU ALREADY KNEW...

Computational Thinking and Programming (1) (LM)
Usability and User Experience (1) (LM)

Basic-intermediate skills
in python programming:
install python libraries,
usage of notebooks
(Jupyter)

Basic knowledge of
Javascript for IU
(modify the DOM,
interactivity).

PYTHON

JAVASCRIPT

WHAT I WISH YOU ALREADY KNEW...

Computational Thinking and Programming (1) (LM)

Create, access and manipulate CSV files with Python and Javascript

Create, access and manipulate JSON files with Python and Javascript

CSV

JSON

WHAT I WISH YOU ALREADY KNEW...

Computational Thinking and Programming (1) (LM)
Introductory seminars
Usability and User Experience (1) (LM)

Good skills in web development (static web pages).

Good skills in usage of version-control interfaces (e.g. sourcetree)

HTML / CSS

GITHUB

WHAT I WISH YOU ALREADY KNEW...

Library, Archive, Museum and Information Science (I.C.) (LM)
Information Technologies and Knowledge Management (I.C.) (LM)

Basics of RDF, syntaxes
(.ttl, RDF/XML), SPARQL,
and OWL.

How to organise knowledge,
how to query/explore data
according to an ontology.

LINKED OPEN DATA

ONTOLOGIES

WHAT I WILL
INTRODUCE YOU

Py Libraries for data exploration.
Jupyter to document your work.

Js Libraries for data visualisation.
Github pages to publish a website.

DATA-RELATED

WEB-RELATED

WHAT I WILL INTRODUCE YOU

Py Libraries for creating RDF data and querying SPARQL endpoints. Short tutorial on SPARQL.

How to frame your work (research questions, exploration, discovery). **Digital storytelling** strategies.

SEMANTIC WEB-RELATED

COMMUNICATION-RELATED

WHAT YOU WILL HAVE TO LEARN BY YOUR OWN

[A short introduction to git and github](#)
[Github guides](#)
[sourcetree GUI to Git](#)

Read online documentation on how version-control systems work.
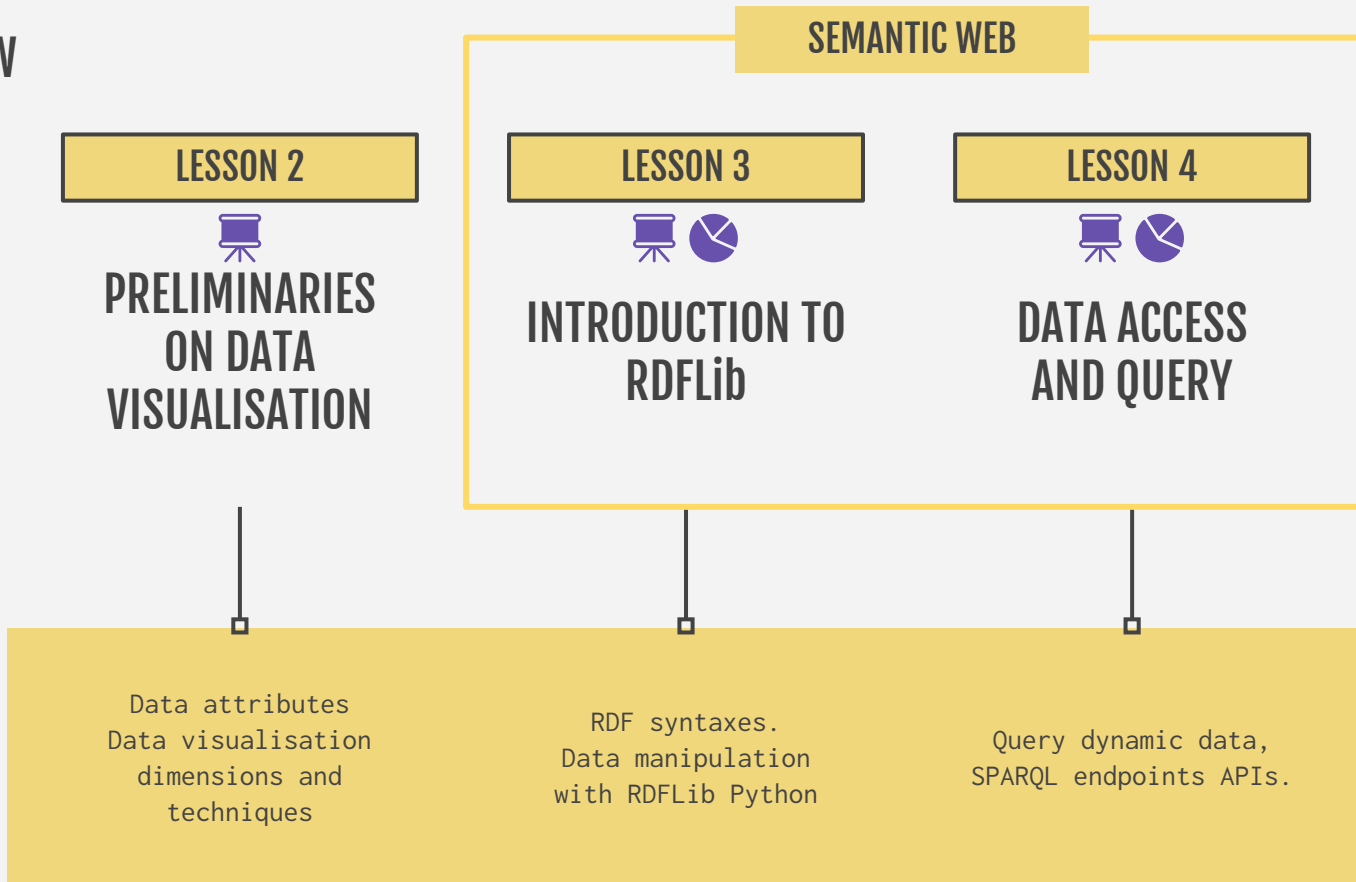
Learn how to use github (via shell or interfaces, e.g. sourcetree)

How to install python and libraries.
How to manipulate data structures (lists and dictionaries).
How to read / write files.

GITHUB

PYTHON BASICS

CLASSES OVERVIEW

SEMANTIC WEB

LESSON 2
🖥
PRELIMINARIES ON DATA VISUALISATION

Data attributes
Data visualisation dimensions and techniques

LESSON 3
🖥 🥧
INTRODUCTION TO RDFLib

RDF syntaxes.
Data manipulation with RDFLib Python

LESSON 4
🖥 🥧
DATA ACCESS AND QUERY

Query dynamic data, SPARQL endpoints APIs.

🖥 lecture

🥧 hands-on

CLASSES OVERVIEW

LESSON 5	LESSON 6	LESSON 7

DATA SENSE MAKING

DATA ANALYSIS WITH PYTHON

DATA VISUALISATION WITH JS

Formulate questions and answers with data visualisation. Introduction to Jupyter.

Data wrangling and analysis

Web development. JS libraries for data visualisation.

lecture

hands-on

CLASSES OVERVIEW

LESSON 8	LESSON 9	LESSON 10

DIGITAL STORYTELLING

SEMINAR

WEB PUBLICATION AND WRAP-UP

lecture

hands-on

Communication strategies and best practices

Federico Zeri Foundation. Art historical linked data
Lecture and tutorial

Binder
Github pages

THE PROJECT

A DATA VISUALISATION PROJECT

You choose a **topic**, find relevant **research questions** that can be answered by visualisations (or that can support domain experts to understand certain phenomena), **analyse** data, produce a **notebook** with your code and a **website** for presenting results.

WHAT YOU NEED

Given a broad domain, pick a problem or aspect that has not been investigated via quantitative methods

THE TOPIC

One or more available datasets, either in Linked Open Data or to be transformed in RDF.

THE DATA

3-people groups to organise the work and share your problems.

A GROUP?

THE TOPIC(S)

History of Art

Artworks provenance
E.g. What are the artworks that <u>travelled</u> most in Europe?

Iconography
E.g. What <u>subjects</u> are mostly represented in artworks of 16th century?

Connoisseurship
E.g. What are the most reliable <u>criteria</u> to justify an artwork attribution?

History of art history

Art historians' relations
E.g. Which countries are historians from and where did they work?

The debate on research topics (artists, periods, movements)
E.g. In which periods the research focused on Mannerism?

Resources
E.g. which types of resources in archival collections represent research topics?

History of Photography of Art

Photographers' relations
E.g. Which photographers worked in Italy in 20th century?

The most photographed artworks
E.g. What art genres are mostly represented in professional photography?

Archives and museums
E.g. Which museums commissioned photographs of artworks?

For whom?

In this course we address a topic that is relevant to scholars, cataloguers, market dealers, museum curators, etc.

SUGGESTIONS
Research questions must be answered by data visualisation (no **general** questions)
The analysis must be tailored to what is relevant to the target, e.g. historians (no **naive** research questions)
The analysis must provide new insights (no **rhetorical** questions)

Your work can be useful to others

Create social value

You create social value by considering the economic, environmental, and social aspects that can impact people's life, increasing their well-being and development.

In Arts, social value is given by the **effects of artistic activities** not taken into account by the market: e.g. increased self-belief, self-empowerment

Gender in Art and Photography history

Representativeness of female or non-binary gender in history
E.g. How influential were <u>female photographers</u> in the 20th century?
E.g. Are female photographers <u>under-represented</u> in photo archives?

Photographers' occupations
E.g. What other <u>occupations</u> had female art historians?

Women and market
E.g. Who <u>buys</u> women's art?

How do I know I picked the correct topic?

Once you defined your research questions, drop an email to marilena.daquino2@unibo.it for feedback.

SPOILER ALERT This will ensure your work is not going to be trashed and can be useful to somebody in the near future (see at the end of this presentation for clarifications)

THE DATA

ARTchives Linked Open Data

http://artchives.fondazionezeri.unibo.it/

ARTchives includes data created by cataloguers of art historical photo archives and reuses data harvested from Wikidata. Data includes information on art historians, archival collections, debated art genres, and keeping institutions.

Zeri Linked Open Data

http://data.fondazionezeri.unibo.it/

Zeri & LODE includes data created by the Zeri Foundation and reuses data harvested from Wikidata, DBpedia, ICONCLASS, AAT Getty, VIAF. Data includes information on artworks and photographs of artworks collected by Federico Zeri, one of the most notable connoisseurs of last century. The dataset is limited to artworks of Modern Art.

External data sources

You can either use ARTchives/ZERI&LODE data only or (**recommended**) integrate data with other relevant sources, both Linked Data or not (e.g. Wikidata, DBpedia, artistorians.info).

If the research questions you have chosen cannot be answered with the two aforementioned datasets **only**, feel free to find external sources that best suit your goals. For instance, you may collect data from other photo archives, e.g. PHAROS (https://vision.artresearch.net/sparql),

If the research questions you have chosen cannot be answered with such datasets **at all**, you can work directly on other sources (e.g. Wikidata)

Wikidata

https://query.wikidata.org/

Wikidata is a general purpose Linked Open Dataset, originally born to represent structured data of Wikipedia (the right-side boxes) in RDF. It includes a variety of information, such as people biographical data, scholarly publications, historical events, and so on.

THE GROUP

GROUPS vs LONELY WOLVES

3 people max.
You need to **justify** your contribution to the project. Grades are individual.

You can work alone, but you need to achieve the same results.
No discounts :)

GROUPS

WOLVES

EVALUATION

WHAT HAPPENS THE DAY OF THE EXAM

it's not an exam on python or javascript!

Live or remote. Everybody can attend by requesting the link to Teams meeting.

While you talk I'll have a look at your source code.

General questions on how you handled the problem and on the **soundness** of your project.

PRESENT YOUR PROJECT

DIG INTO THE CODE

ANSWER QUESTIONS

WHAT YOU NEED TO HAVE
BY THE DAY <u>BEFORE</u> THE EXAM

A Jupyter notebook **published on the web** including documented workflow to create and analyse your data

JUPYTER NOTEBOOK

A static web page or a web application **published on the web** showing results of your data analysis

WEB PAGE

A presentation of your work according to the template I give you. With slides or not.

15min PRESENTATION

PREPARE YOUR MATERIALS: THE JUPYTER NOTEBOOK

1. Install **Jupyter** [1] on your laptop
2. Create your <projectname> notebook
3. Describe your project aims in a short **Abstract** section.
4. Prepare (extract, integrate, clean) the data you want to visualise and save them into a **CSV or JSON file**.
5. For every function you define and for each step in the data analysis, add an explanation in the notebook in **Markdown** [2] syntax.
6. [If you are a group] Merge your notebooks into a **single notebook**.

[1] https://jupyter.org/
[2] https://www.markdownguide.org/

PREPARE YOUR MATERIALS: THE JUPYTER NOTEBOOK

1. Create an account on **GitHub** (individual)
2. Create a repository for the project (collective or individual)
3. Upload the jupyter notebook and the CSV/JSON data to the repository.
4. Connect your repository to **Binder** [1] to make the notebook interactive on the web.
5. Include the binder link into your README file in the repository.

DONE!

[1] https://mybinder.org/

PREPARE YOUR MATERIALS:
THE WEB PAGE

1. Create a **index.html** page
2. Create a **CSS** stylesheet for the webpage
3. Create a **Javascript** file
4. Access the data previously created (CSV/JSON) from the JS file
5. Add style to your data (CSS/JS).
6. Upload the website files to the repo.
7. Use **github pages** to publish your website.

DONE!

[1] https://github.com/marilenadaquino/epds

PREPARE YOUR MATERIALS:
LINK YOUR PROJECT

1. Open the repository of the course [1]
2. Open or comment on an **issue** called "Exam DD/MM/YYYY" (or similar) with:
   a. Project title
   b. Website URL
   c. Repository URL (if not github)
   d. People involved

DONE!

[1] https://github.com/marilenadaquino/epds

PREPARE YOUR MATERIALS: RESPONSIBILITY AND LICENSE

IF COLLABORATIVE PROJECT
Please MAKE CLEAR who was responsible for what aspect of the project (e.g. data access and cleaning, data analysis, data visualisation, web development, graphics, communication strategy and so on) during the presentation

BOTH INDIVIDUAL/COLLABORATIVE PROJECTS
Specify your names and tasks:
- in the README file of the repository
- on the webpage (e.g. in the footer, in a dedicated page called *Credits*)

LICENSE FOR THE REUSE OF DATA
In the README file of the repository specify:
- The license of data you reused (look at their websites)
- The license of your derivative data (please, use either CC0 or CC-BY licenses to allow future reuse)

PREPARE YOUR MATERIALS:
DO NOT PANIC

no need to do everything from scratch!

*You can either create your own static web page with your styles or you can **reuse existing projects or tools** (e.g. CMS, HTML templates)*

*You don't need to style from scratch the visualisation, you can (definitely) **reuse existing libraries**.*

no need to use github to publish your website!

Github is a free-of-charge solution to publish a project website and to handle the code in the same environment. Feel free to use other solutions for publishing the website! However, it is mandatory for publishing the notebook and the data.

but especially, no need to panic!

We will set up all the pieces of your project (install Jupyter, use libraries, create an account on github, create and publish a webpage) in dedicated hands-on classes

PREPARE YOUR MATERIALS: THE PRESENTATION

If you present with a slide presentation, please make it 10 slides max.
No need to share it before the exam day.

Use the following **template** to prepare the speech.

Title
1. Background
2. Goals
3. Research questions
4. Data preparation and data analysis
5. Data visualisations selected and reasons
6. Data communication strategies
7. Summary of results

AN EXAMPLE:
THE PRESENTATION

Title: **Trends in the study of artistic periods**

Background. Artistic periods are differently studied by art historians over time. An artistic period may be overlooked at a certain time, possibly due to a decreased interest (market, research discoveries, exhibitions) towards some artist or genre.

Goals. Discover trends of interest towards artistic periods by looking at the dates of activity of art historians (and their archival collections) that studied a certain period.

Research questions.
- When there is a research interest towards a certain artistic period?
- How the interest evolves over time?
- Which artistic periods show a significant trend over time?

AN EXAMPLE:
THE PRESENTATION

Data preparation and data analysis.
We studied ARTchives and Wikidata data models.
Data about archival collections, art historians' activity dates, and related artistic periods are collected from ARTchives. Descriptions and dates of artistic periods are collected from Wikidata.
We query ARTchives and Wikidata SPARQL endpoints, we reconcile the data, we prune duplicate concepts (e.g. "Baroque" and "Baroque art").
We perform some preliminary analysis to understand the distribution of periods over the archival collections.
We analyse the trend of artistic periods as subject of art historians' collections over time.

Data visualisations selected and reasons. We show trends in a line chart having on the x axis the dates of art historians activities (corresponding to the time range of their collection) and on the y axis the artistic periods. So doing a user can see all trends at the same time and figure if there are correlations in trends.

AN EXAMPLE:
THE PRESENTATION

Data communication strategies. We first show preliminary exploratory visualisations about the distribution of periods as subjects addressed by art historians so as to demonstrate the validity and representativity of results. We provide brief descriptions of artistic periods for whom may not be acquainted with art history. We show results of our investigation as an interactive line chart where users can select/remove the period to be shown in the graph.

Summary of results. Most significant results show that artistic period XX received lots of attention in the 19XXs, while period YY shows a significant loss of attention in the same period. Notably, period XX and ZZ have similar trend over time.

[TO EXPLAIN WHY THIS HAPPENS IS NOT UP TO YOU - if you are not an art historian]

What I value most

Correctness and efficiency of the **SPARQL queries**

Correctness of **final CSV/JSON data** (after data cleansing)

Data integration (reconciliation and enrichment) with external sources

Soundness and usefulness of research questions

Choice of the most **representative** visualisations

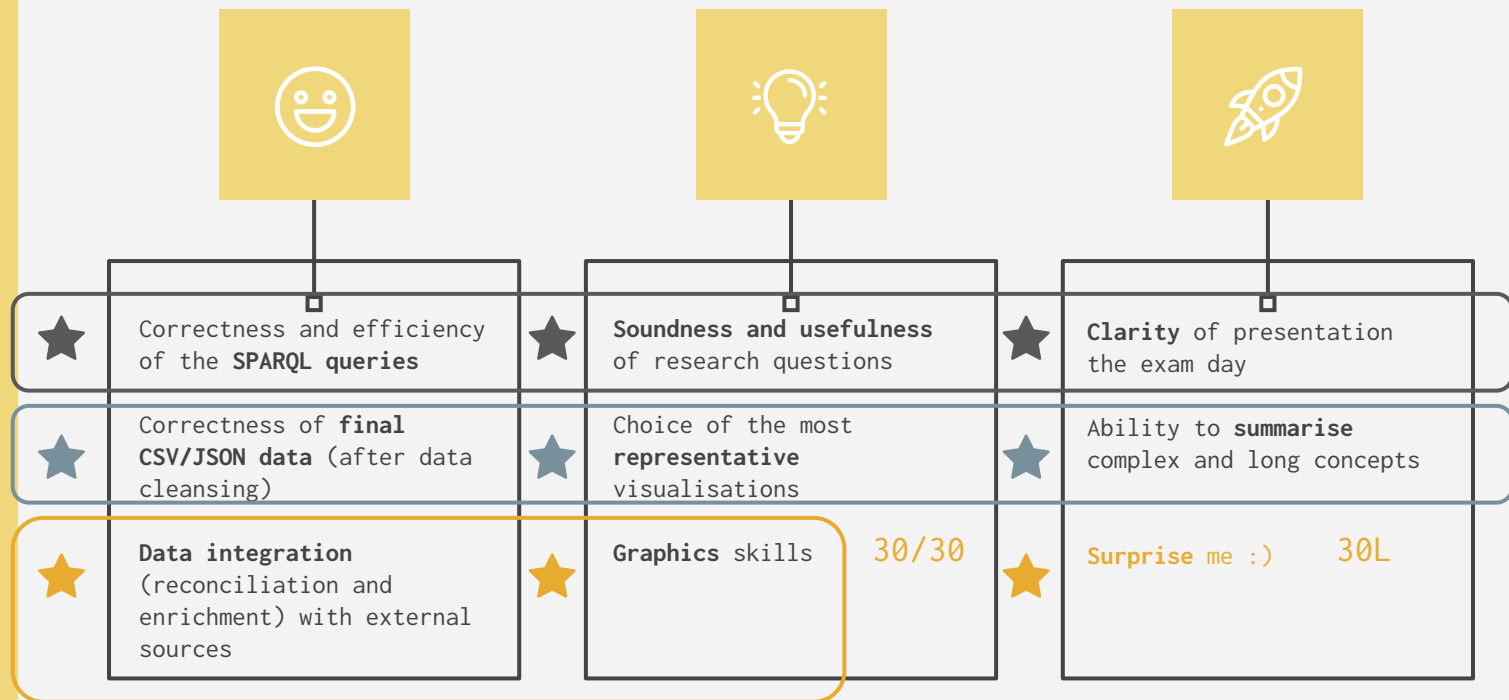**Graphics** skills

Clarity of presentation during the exam

Ability to **summarise** complex and long concepts

Surprise me :)

What I value most

☺	💡	🚀

⭐ Correctness and efficiency of the **SPARQL queries**

⭐ **Soundness and usefulness** of research questions

⭐ **Clarity** of presentation the exam day

24/30

⭐ Correctness of **final CSV/JSON data** (after data cleansing)

⭐ Choice of the most **representative** visualisations

⭐ Ability to **summarise** complex and long concepts

28/30

⭐ **Data integration** (reconciliation and enrichment) with external sources

⭐ **Graphics** skills

30/30

⭐ **Surprise** me :)

30L

Explore ARTchives

Explore art historians' archival collections th

Discover archival collections that are relate

historical periods, artists, art historian

organisations.

A SMALL REWARD

The most insightful and well-designed visualisations will be published/linked to the website.

ARTISTIC PERIODS

Temporal data are collected from Wikidata

Middle Ages

Explore archival collections through

FONDAZIONE
FEDERICO ZERI
UNIVERSITÀ DI BOLOGNA

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
A.D. 1088

WORK FOR THE ZERI FOUNDATION

The Zeri foundation is working on statistical analysis of their collections. Projects may address aspects that can be generalised to their data and become part of their new application.

YET ANOTHER PERSPECTIVE

Co-authoring a scholarly publication
dedicated to quantitative history
(in a conference or journal)

LET'S TRY TO GET INTO THE SPIRIT (1)

1. Read this article (20min) on Medium
Daniel Bourke, A gentle introduction to Exploratory Data Analysis. Medium 2019.

Take away message (If you run out of time)
The preliminary questions when exploring a new dataset:

1. What question(s) are you trying to solve (or prove wrong)?
2. What kind of data do you have and how do you treat different types?
3. What's missing from the data and how do you deal with it?
4. Where are the outliers and why should you care about them?
5. How can you add, change or remove features to get more out of your data?

LET'S TRY TO GET INTO THE SPIRIT (2)

2. Explore ARTchives (30 min) and try to (roughly) answer the questions in the prior slide
Understand what types of data are there by navigating through some cataloguing records.

HINT 1: Use the question shown in the presentation example
Trends in the study of artistic periods

HINT 2: Here a preliminary visualisation that helps you
http://artchives.fondazionezeri.unibo.it/contents

3. Fill in the questionnaire (**by the next class**) with your answers
https://forms.gle/gJ1NFrBRzpuDbssg9

WHY A QUESTIONNAIRE?

A few questionnaires will be given to you over time. These are primarily meant to give the teacher **an estimate of your general understanding**.

Results of the tests will not contribute to the final grade, hence these are not mandatory, you should not be afraid to answer wrong (this won't be used against you), and you may even decide to use a pseudonym if you don't want me to know your identity (just be consistent please!).

WHY A QUESTIONNAIRE?

However...

Remember that the final presentation lasts 15 minutes, and it's the first moment I get to know you closely (which can be either good or bad).

If I knew you had **good** results in the questionnaires and you gave a **good** presentation this may encourage me to give you the maximum grade.	If I knew you had **good** results in the questionnaires but you gave a **bad** presentation, I'd take into account your constant effort and you'd not be disadvantaged.
If I knew you had **bad** results in the questionnaires but you gave a **good** presentation this would highly impress me and I'd tend to be more generous.	If I knew you had **bad** results in the questionnaires and you gave a **bad** presentation, well...at least you tried!

Take your chance!

THANKS

Does anyone have any questions?

marilena.daquino2@unibo.it
github