OPEN DATA EXAM (June 3rd, 2014)

Name: .....

PREGUNTA 1. [4p]

The authorities of Gotham City decide to create a BI tool to fight crime purely based on tons of Open Data information made available by the criminals in Web Forums, Social Networks (such as Twitter and Facebook) and also public crime statistics made available as RDF triples by the town council (crime rate, kind of crime, etc.). The authorities are mainly interested in crossing data and run Machine Learning algorithms to identify trends or unexpected behaviours.

Sketch the main modules that such system should implement. Think of a functional architecture similar to the one you had to produce for your course project (i.e., conceptual modules without needing to link them to specific technologies but functionalities).

Consider now the following kind of sources: social networks (e.g., Twitter, Facebook), web forums (i.e., traditional web interfaces) and RDF triples about official crime statistics opened by the city council. For each of the three kinds specify how would you propose to extract data from them (you may consider any of the technologies seen in the course).

- Social Networks:

- Web Forums:

- Linked Open Data:

PREGUNTA 2. [4p]

Justify the metadata repository you implemented in your course project. Specifically, clearly discuss and justify the following points:

- What metadata artefacts (group them in categories according to the purpose they are being used for) were defined for your course project?

- For each category, what is its main role? Briefly explain how they enable automation of different aspects of the system.

PREGUNTA 3. [2p]

Your project course global schema is the one accessed by the end-user and thus, making transparent to the end-user the existence of several data sources. Explain the integration rules you used in your project to transform the data source instances into global schema instances.

Suppose now you decide to implement your global schema as an ontology. Following the linked data spirit, you choose *RDF Schema* as ontology language. When using inference / reasoning on your schema, will the open world assumption hold? Justify your answer.