mappa

metodologie applicate alla predittività del potenziale archeologico

www.mappaproject.org

Ha(r)ckeology

Gabriele Gattiglia

REGIONE TOSCANA

FAS Fondo Aree Sottoutilizzate 2007-2013

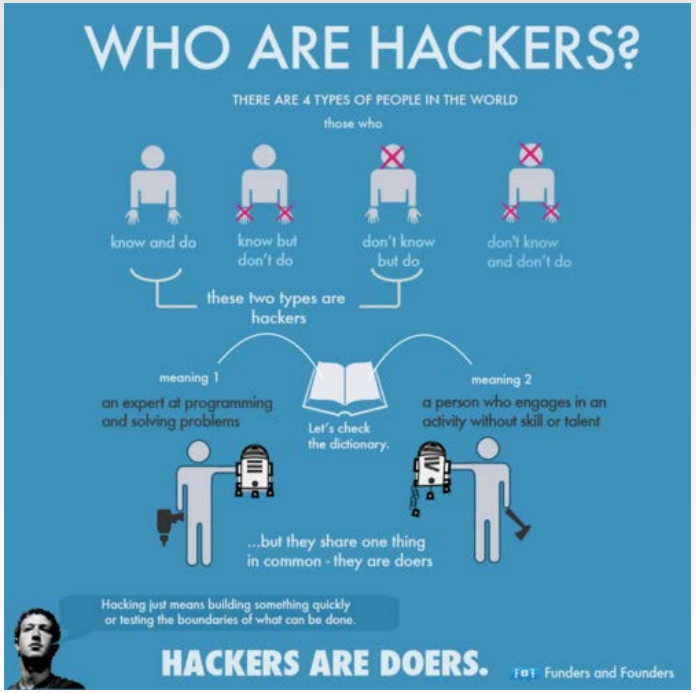
UNIVERSITÀ DI PISA
Dipartimento di Scienze Archeologiche
Dipartimento di Scienze della Terra
Dipartimento di Matematica

we want to encourage a civic hacking approach to archaeology, an open data approach that we define as ha(r)ckeology.

Hacking is a tricky term which it's difficult to find a single widely accepted definition for. To most of the population, hacking is still associated solely with the acts of breaking into security systems. To those near the technology world, hacking means attempting to solve problems more quickly or creatively than before, and it's about using new ideas and approaches to improve the *status quo*. **Hacker is not someone who is able to subvert computer security; if doing so for malicious purposes, the person can also be called a cracker!**

mappa
metodologie applicate alla predittività del potenziale archeologico

In the context of ha(r)ckeology, we're clearly more interested in the problem-solving definition of hacking. Using Mark Zuckerberg's definition: hacking just means building something quickly or testing the boundaries of what can be done. A definition with no explicit mention of technology: hacking is conceived more as a process than as a specific toolset.



mappa
metodologie applicate alla predittività del potenziale archeologico

www.mappaproject.org

Therefore, ha(r)ckeology is the act, conducted by archaeologists, of quickly improving the processes and systems of archaeology with new tools or approaches, or more simply ha(r)ckeology means archaeologists working together quickly and creatively to improve archaeology. For reaching such a goal, the first step is to educate a new generation of archaeologists, a sharing generation able to work with a trowel, and to share and manipulate data, a generation that is aware that archaeological data must be open because they are public, they are expensive to produce, and they must be reused.

mappa
metodologie applicate
alla predittività
del potenziale archeologico

www.mappaproject.org

Let’s start…

Tabular data

mappa
metodologie applicate alla predittività del potenziale archeologico

www.mappaproject.org

- **Tabular data….** ***The Good, the Bad and the Ugly***

mappa
metodologie applicate alla predittività del potenziale archeologico

pdf portable document format

example

Web

.xls

.csv comma separated value

example

.json JavaScript Obiect Notation

example

mappa
metodologie applicate
alla predittività
del potenziale archeologico

www.mappaproject.org

OpenRefine

is a standalone open source desktop application for data cleanup and transformation to other formats, the activity known as data wrangling. It is similar to spreadsheet applications (and can work with spreadsheet file formats), however, it behaves more like a database

http://openrefine.org/

TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML

mappa
metodologie applicate alla predittività del potenziale archeologico

data wrangling is the process of manually converting or mapping data from one "raw" form into another format that allows for more convenient consumption of the data with the help of semi-automated tools.

The main step of this process are:

- extracting the data in a raw form from the data source,
- "munging" the raw data using algorithms (e.g. sorting) or parsing the data into predefined data structures
- depositing the resulting content into a data sink for storage and future use

In the scientific research context, a **data wrangler** is the person responsible for gathering and organizing disparate data sets collected by many different investigators

http://en.wikipedia.org/wiki/Data_wrangling

mappa
metodologie applicate alla predittività del potenziale archeologico

www.mappaproject.org

GitHub is a web-based Git repository hosting service.
Git working directory is a full-fledged repository with complete history and full version-tracking capabilities
Unlike Git, which is strictly a command-line tool, GitHub provides a web-based graphical interface and desktop as well as mobile integration. It also provides access control and several collaboration features such as wikis, task management, and bug tracking and feature requests for every project.

https://github.com/ggattiglia/Ha-r-ckerology/issues/1

How can we extract data, if tables are locked inside a pdf file?

mappa
metodologie applicate alla predittività del potenziale archeologico

KEEP CALM AND LET'S TRY AGAIN

Tabula allows us to extract data locked inside a pdf file into a CSV using a simple, easy-to-use interface.

mappa
metodologie applicate alla predittività del potenziale archeologico

Geocoding is the process of enriching a description of a location, most typically a postal address or place name, with geographic coordinates.

https://developers.google.com/maps/documentation/geocoding/index

http://www.geonames.org/export/web-services.html

mappa
metodologie applicate alla predittività del potenziale archeologico

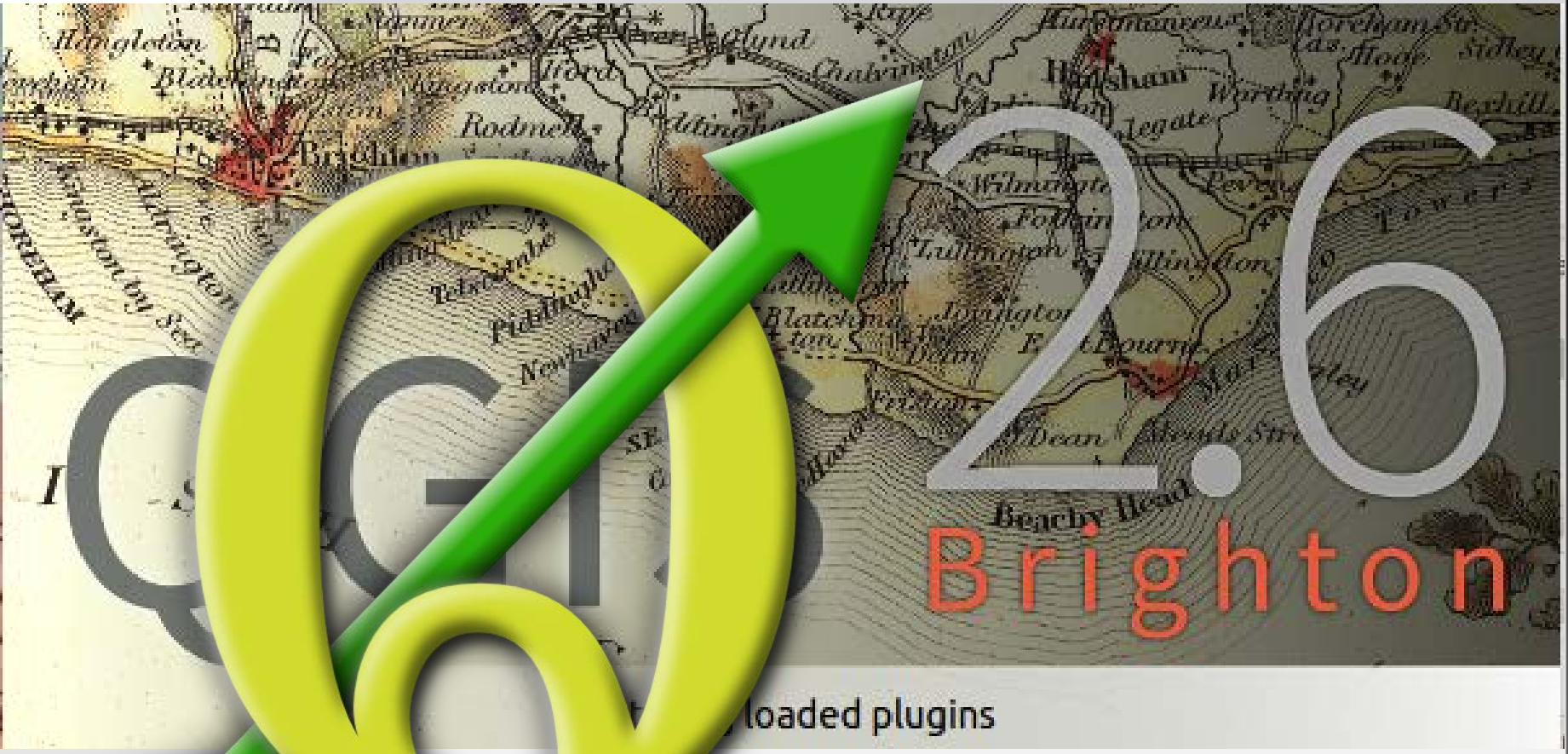
KEEP CALM and WEB SCRAPING

https://import.io/

Web scraping (web harvesting or web data extraction) is a computer software technique of extracting information from websites. Usually, such software programs simulate human exploration of the World Wide Web by either implementing low-level Hypertext Transfer Protocol (HTTP), or embedding a fully-fledged web browser, such as Internet Explorer or Mozilla Firefox.

http://en.wikipedia.org/wiki/Category:Archaeological_sites_in_Iraq

mappa
metodologie applicate alla predittività del potenziale archeologico

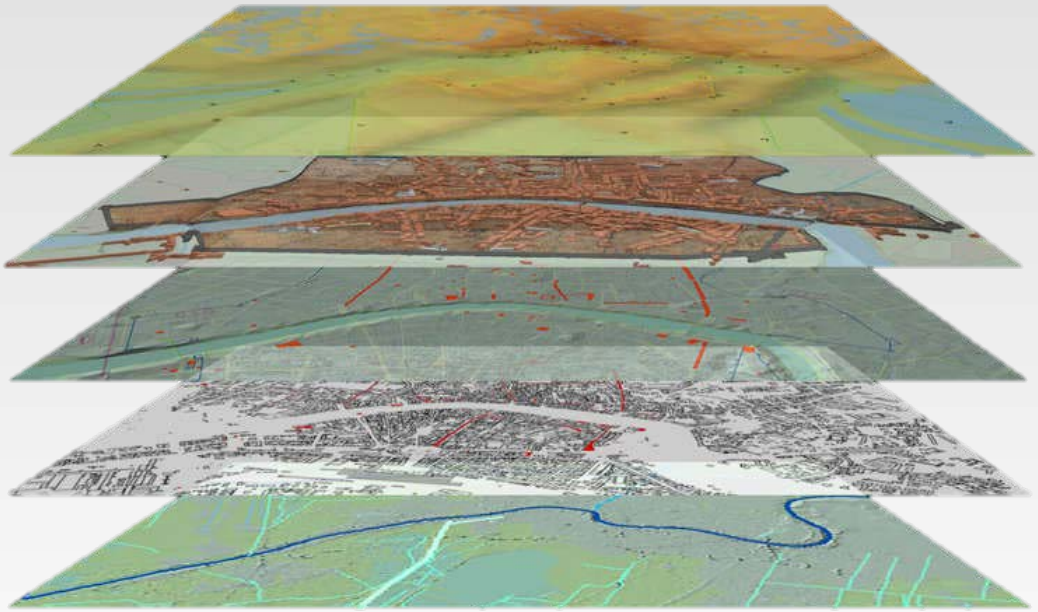


mappa
metodologie applicate
alla predittività
del potenziale archeologico

www.mappaproject.org

2002, Wheatley and Gillings: «Contrary to popular mythology, contemporary professional *archaeologists* may spend *more* time *using GIS than a trowel*»

2014, CAA Session: Is GIS the new trowel?



mappa
metodologie applicate alla predittività del potenziale archeologico

www.mappaproject.org

mappa
metodologie applicate
alla predittività
del potenziale archeologico

www.mappaproject.org