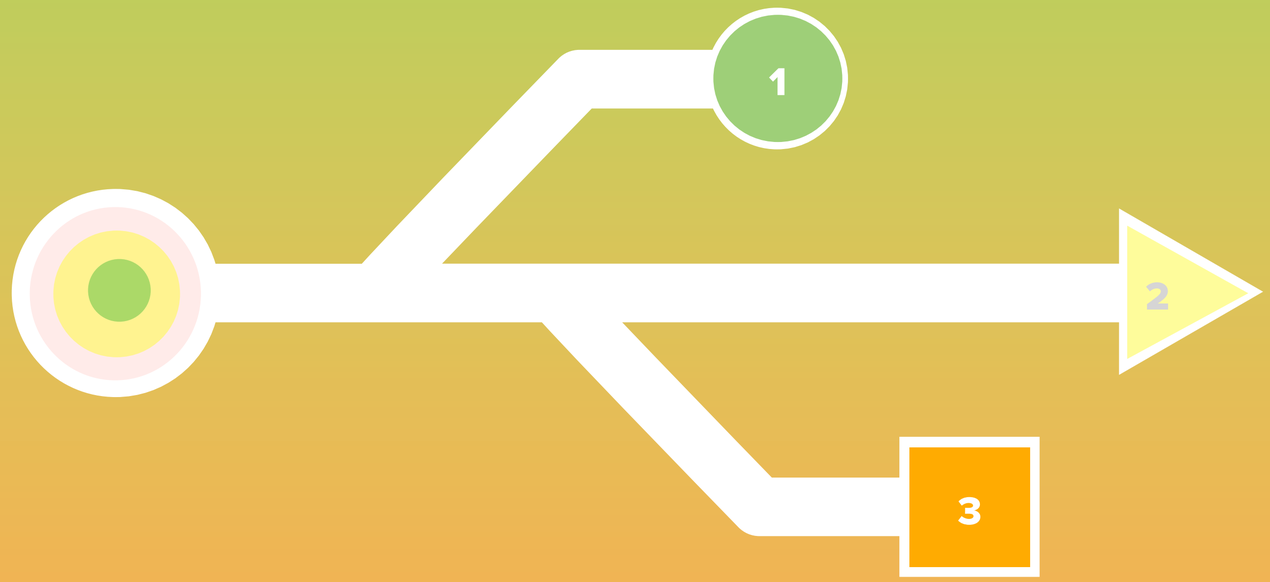
(INFORMATICA PER LE)
DIGITAL HUMANITIES

LECTURE 2

FEBRUARY 24 2023


1
2
3

DH HISTORY

1946

Father Roberto Busa

In 1946 Father Roberto Busa had the revolutionary idea of using computers to study texts, in particular the collected works of St. Thomas Aquinas; and, in 1949, during a trip to New York, he had the chance to present his idea to Thomas Watson, Sr., founder of the IBM Corporation, who decided to support his project.

Humanities

Optical Character Recognition (OCR)
1950s

1949 . FATHER ROBERTO BUSA INDEX TOMISTICUS

Humaniti...
Father R...
Interactive...
Optical Ch...
Index Tho...
Situation...
Dartmouth...
GIS
Labyrinth ...
The Mediu...
Web 2.0
First two...
Creative C...
Creation...
Wikipedia
Creation...
Text Enc...
Creation o...
Google Bo...

1810 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010 2020 2030 2040 2050 2060 2070 2080

Timeline JS

1980

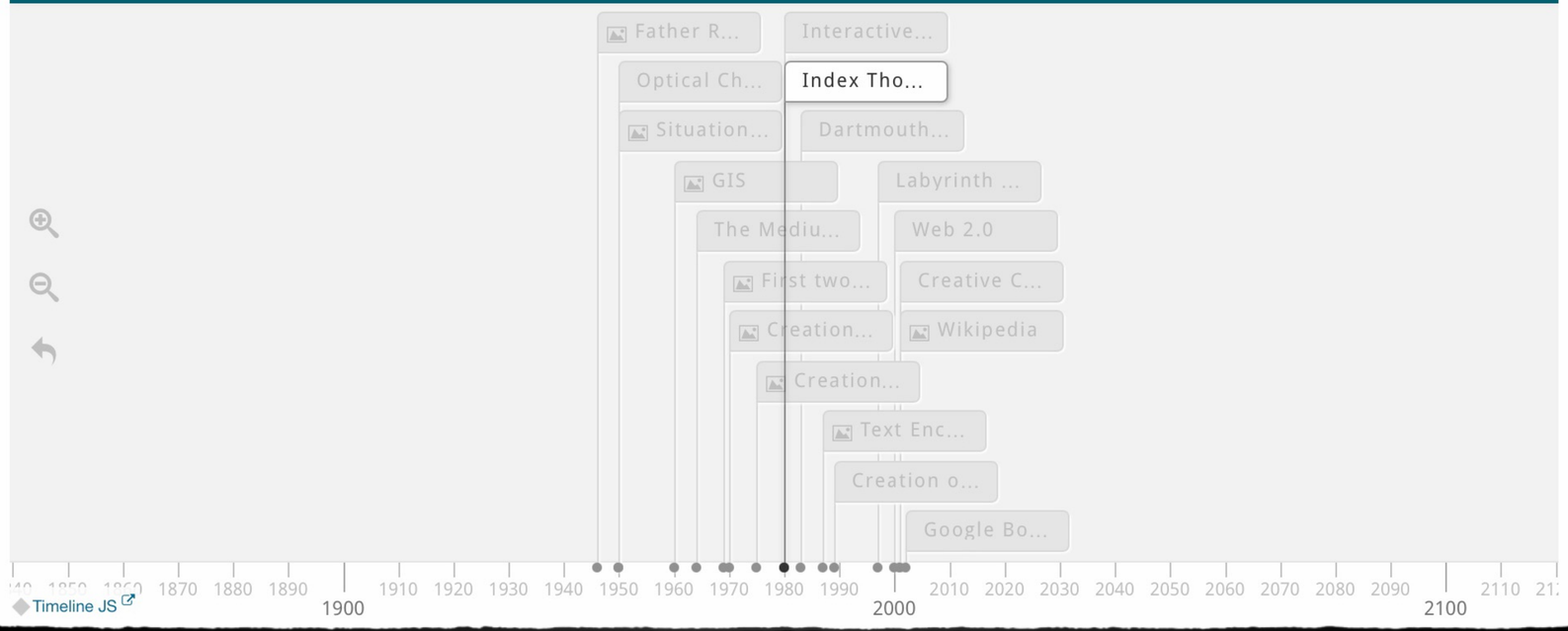
Index Thomisticus

In 1980, after thirty years' work, the printed edition of 56 encyclopedic volumes of the "Index Thomisticus" saw the light of day, an imposing work which gathers the entire production of St. Thomas Aquinas in a format readable and manageable by computer using the methodology developed by Father Busa.

Interactive Cinema Group at MIT

Dartmouth Dante Project
1982-1988

1980 . PRINTED EDITION INDEX TOMISTICUS



1982-1988

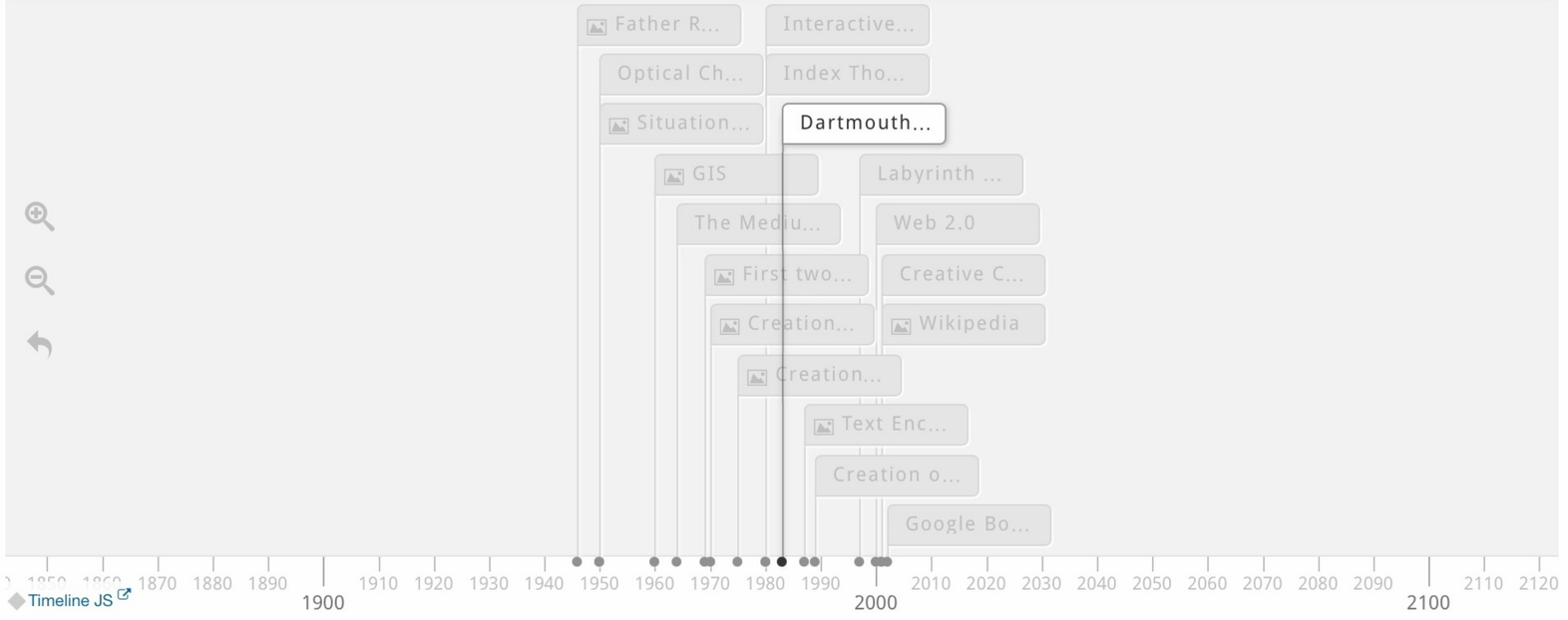
Dartmouth Dante Project

This project combines modern information technology with 75+ commentaries on Dante's Divine Comedy - the Commedia, to create a searchable full-text database.

Index Thomisticus

Text Encoding Initiative (TEI)
1987

1982 - 1988 . DARTHMOUTH DANTE PROJECT



2002

Google Books

A small group of Googlers officially launches the secret “books” project. They begin talking to experts about the challenges ahead, starting with a simple but crucial question: how long would it take to digitally scan every book in the world? After learning that it would take an estimated 1,000 years to scan University of Michigan Library's 7 million volumes, Larry Page (Google's Co-Founder) tells university president Mary Sue Coleman he believes Google can help make it happen in six.

Wikipedia

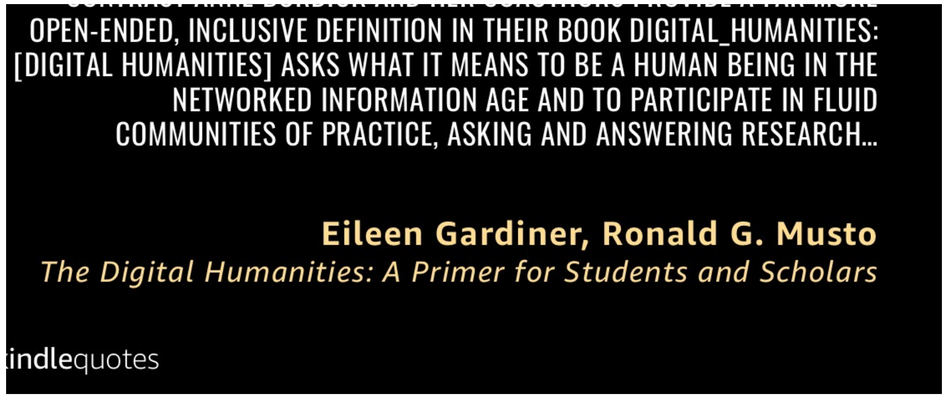
2002 . GOOGLE BOOKS

DH METHOD

DIGITAL SCHOLARSHIP

HUMANITIES TO DIGITAL HUMANITIES

WITH



P.3

THE INTERSECTION OF THE HUMANITIES AND THE DIGITAL CREATED AN ENVIRONMENT IN WHICH THE HUMANITIES BECAME SUBJECT TO NEW APPROACHES THAT RAISED ISSUES ABOUT THE NATURE OF THE HUMANITIES WHILE ALSO OPENING UP NEW RESEARCH METHODS.

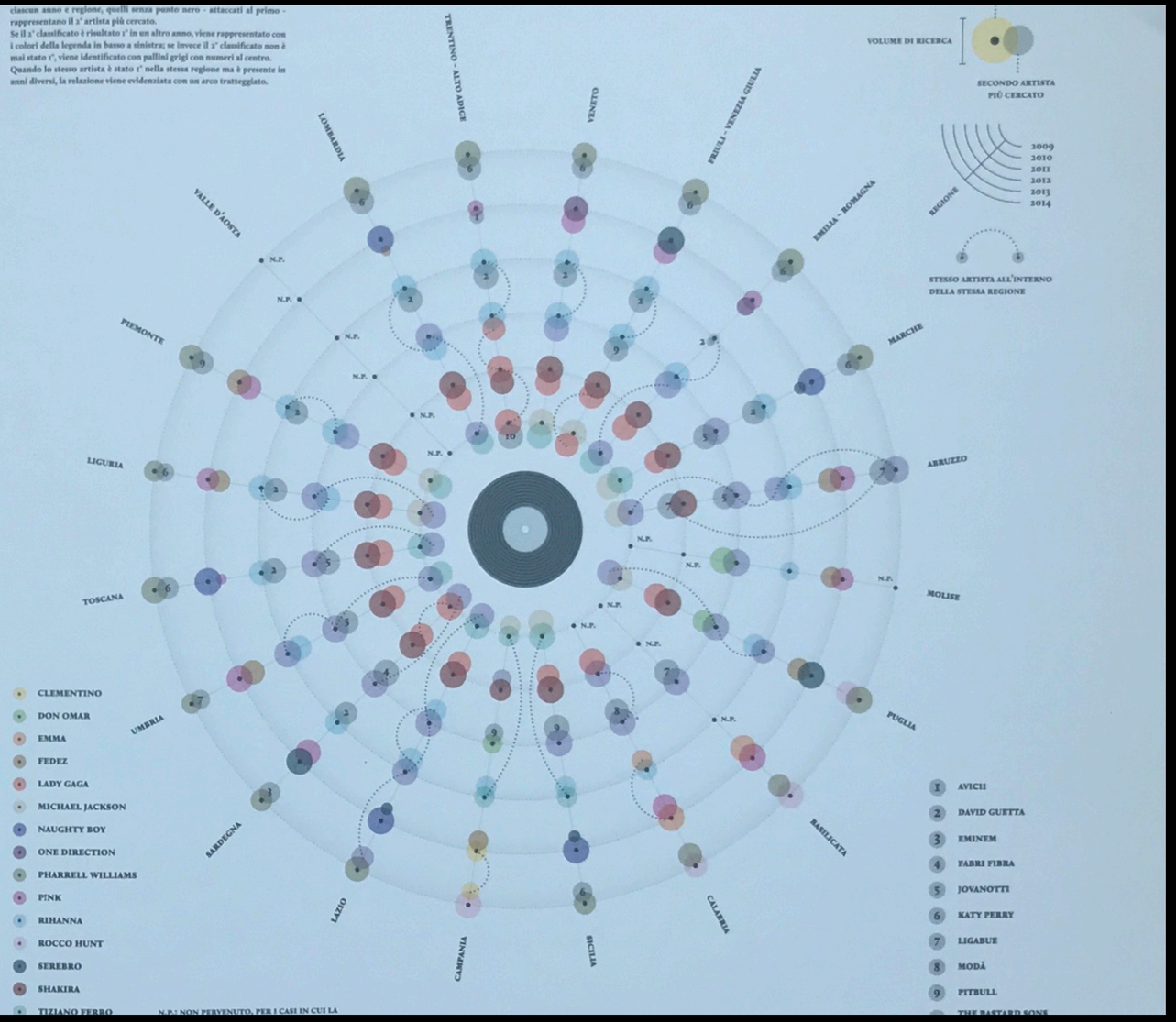
THE ARRAY OF PLATFORMS, APPLICATIONS, TECHNIQUES AND TOOLS, ALL DEVELOPED UNDER THE RUBRIC OF “DIGITAL,” HAVE BEEN DRAMATICALLY CHANGING THE WAY THAT HUMANISTS WORK, HOW THEY DO RESEARCH, GATHER INFORMATION, ORGANISE, ANALYSE AND INTERPRET IT AND DISSEMINATE FINDINGS. HOW DOES THE DIGITAL AFFECT THIS BASIC WORK?

LOOK AND SEE FROM AFAR AND UP CLOSE

ANS ARCHIVIO DI NUOVA SCRITTURA

HTTP://WWW.VERBOVISUALEVIRTUALE.ORG

La visualizzazione è a cura di Michele Mauri, dottorando in Design al Politecnico di Milano che svolge attività di ricerca presso il DensityDesign Lab, e Valerio Pellegrini, information designer e illustratore. Il grafico è stato pubblicato su «la Lettura» il 20 luglio 2014.



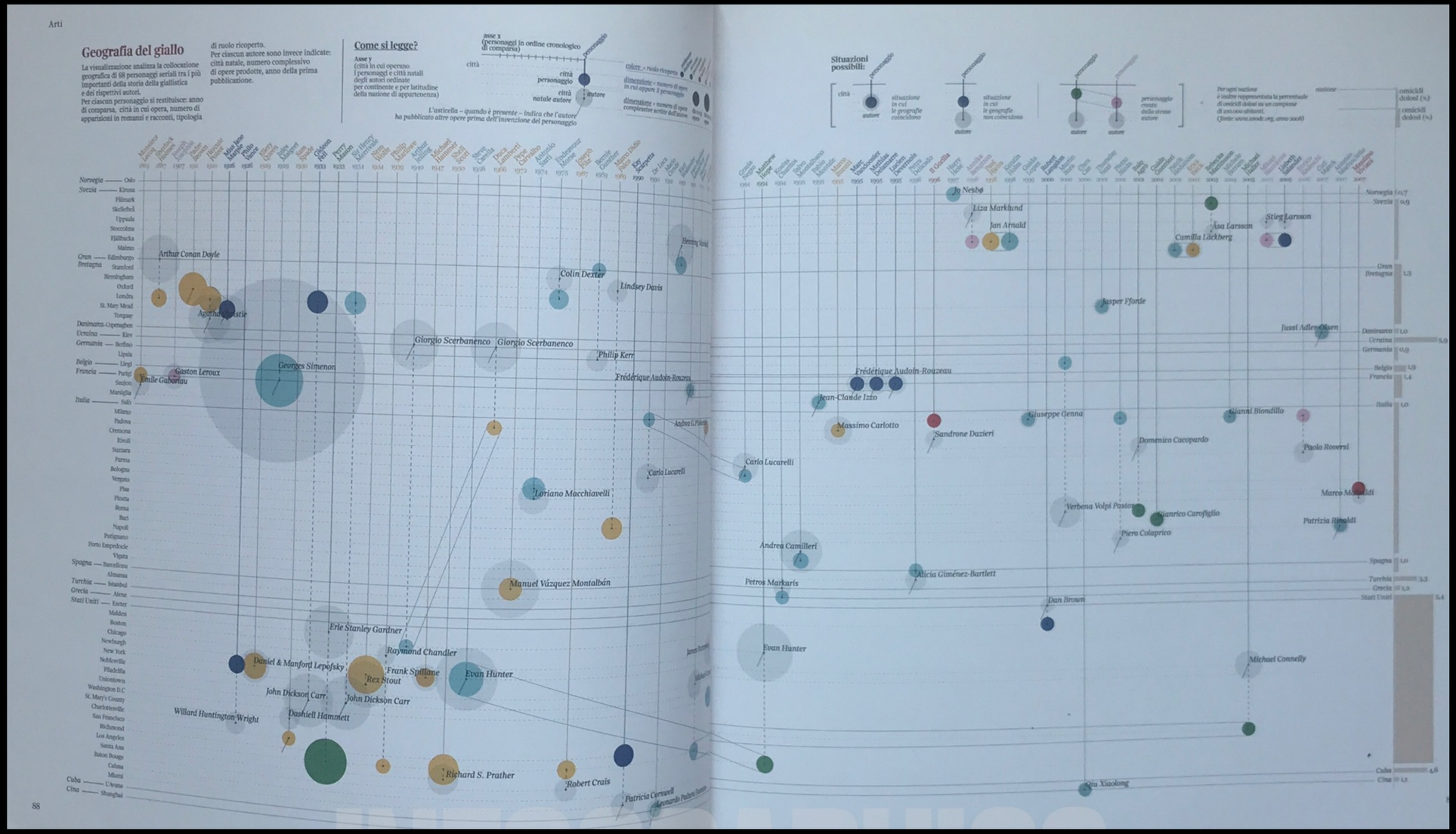
LE MAPPPE DEL SAPERE.VISUAL DATA DI ARTI, NUOVI LINGUAGGI, DIRITTI.L-INFOGRAFICA RIDISEGNA LE CONOSCENZE

EDITED G.COLIN, A.TROIANO

RIZZOLI, CORRIERE DELLA SERA

2014

INFOGRAPHICS



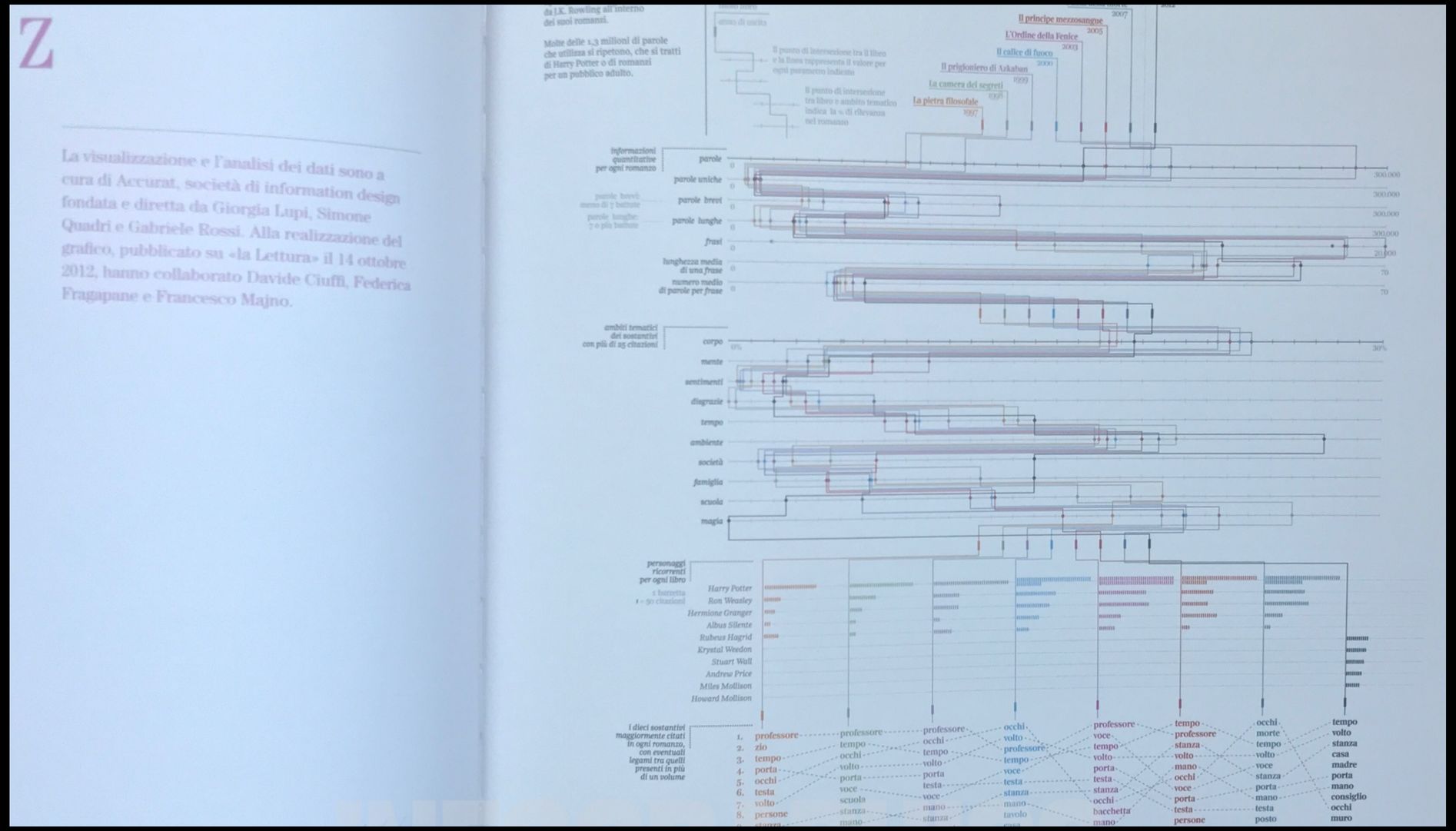
LE MAPPPE DEL SAPERE.VISUAL DATA DI ARTI, NUOVI LINGUAGGI, DIRITTI.L-INFOGRAFICA RIDISEGNA LE CONOSCENZE

EDITED G.COLIN, A.TROIANO

RIZZOLI, CORRIERE DELLA SERA

2014

INFOGRAPHICS



LE MAPPPE DEL SAPERE.VISUAL DATA DI ARTI, NUOVI LINGUAGGI, DIRITTI.L-INFOGRAFICA RIDISEGNA LE CONOSCENZE

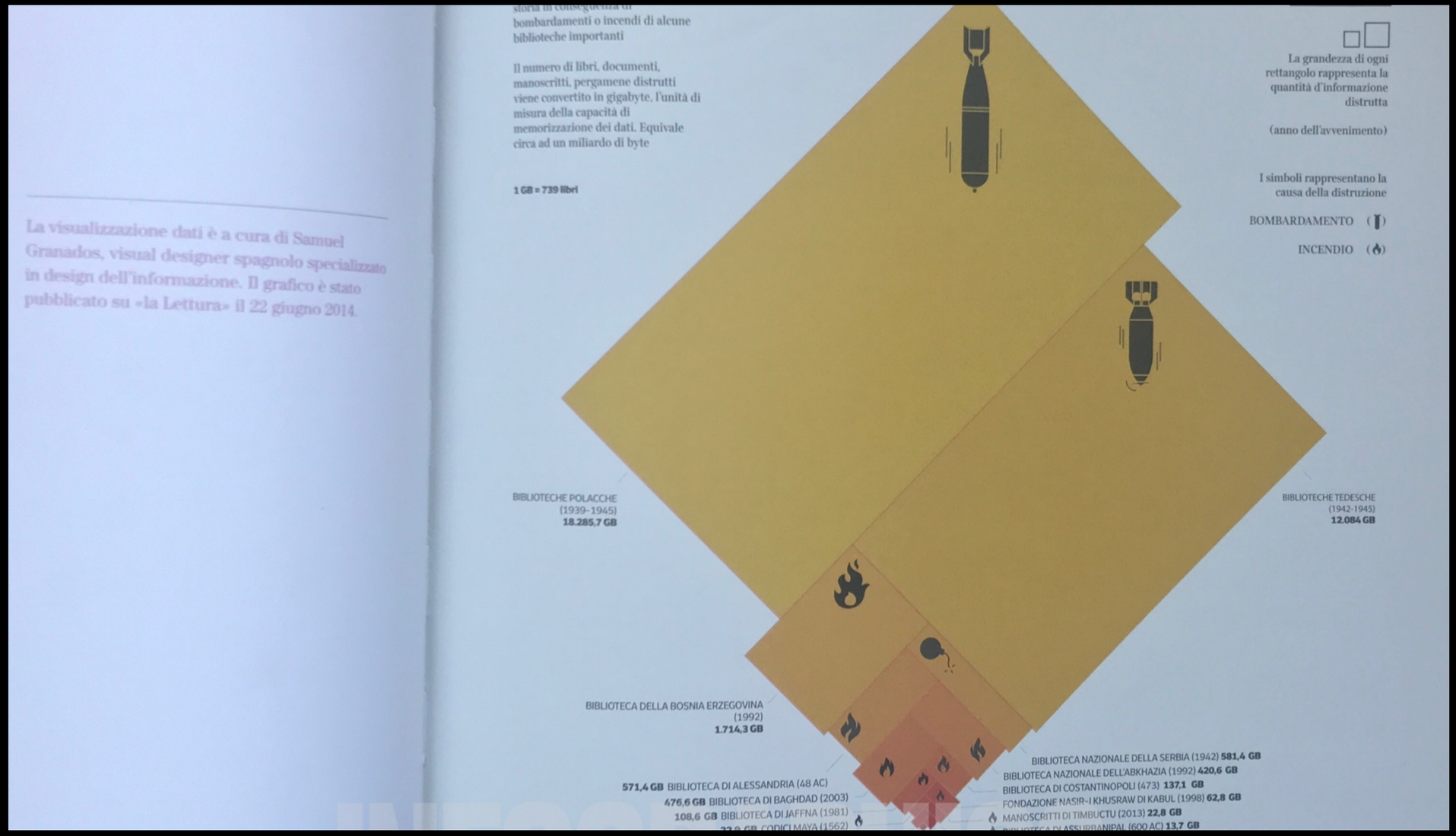
EDITED G.COLIN, A.TROIANO

RIZZOLI, CORRIERE DELLA SERA

2014

INFOGRAPHICS

La visualizzazione dati è a cura di Samuel Granados, visual designer spagnolo specializzato in design dell'informazione. Il grafico è stato pubblicato su «la Lettura» il 22 giugno 2014.



LE MAPPPE DEL SAPERE.VISUAL DATA DI ARTI, NUOVI LINGUAGGI, DIRITTI.L-INFOGRAFICA RIDISEGNA LE CONOSCENZE

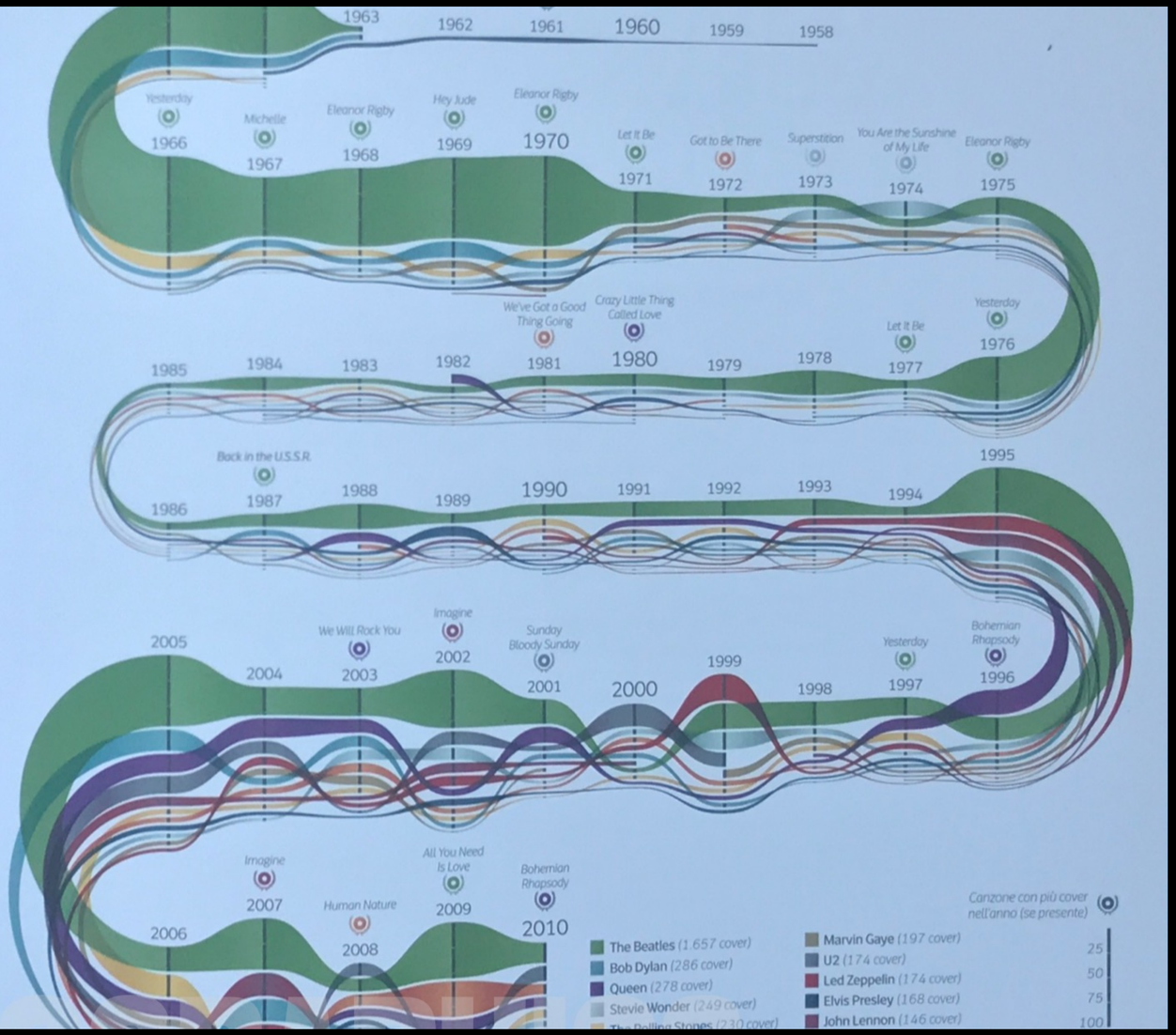
EDITED G.COLIN, A.TROIANO

RIZZOLI, CORRIERE DELLA SERA

2014

INFOGRAPHICS

La visualizzazione dati è a cura del DensityDesign Lab del Politecnico di Milano, Dipartimento di Design, guidato da Paolo Ciuccarelli. Il grafico, pubblicato su «la Lettura» l'8 gennaio 2012, è stato realizzato da Michele Mauri e ha vinto il primo premio nella sezione Infographic/Infodesign alla prima edizione del Kantar Information Is Beautiful Award, tenutasi a Londra nel settembre 2012. Il premio è stato ideato da David McCandless, scrittore, data journalist e information designer, autore del libro *Information is Beautiful* (HarperCollins 2009, edizione italiana BurExtra 2011) sulla visualizzazione dell'informazione.



LE MAPPPE DEL SAPERE.VISUAL DATA DI ARTI, NUOVI LINGUAGGI, DIRITTI.L-INFOGRAFICA RIDISEGNA LE CONOSCENZE

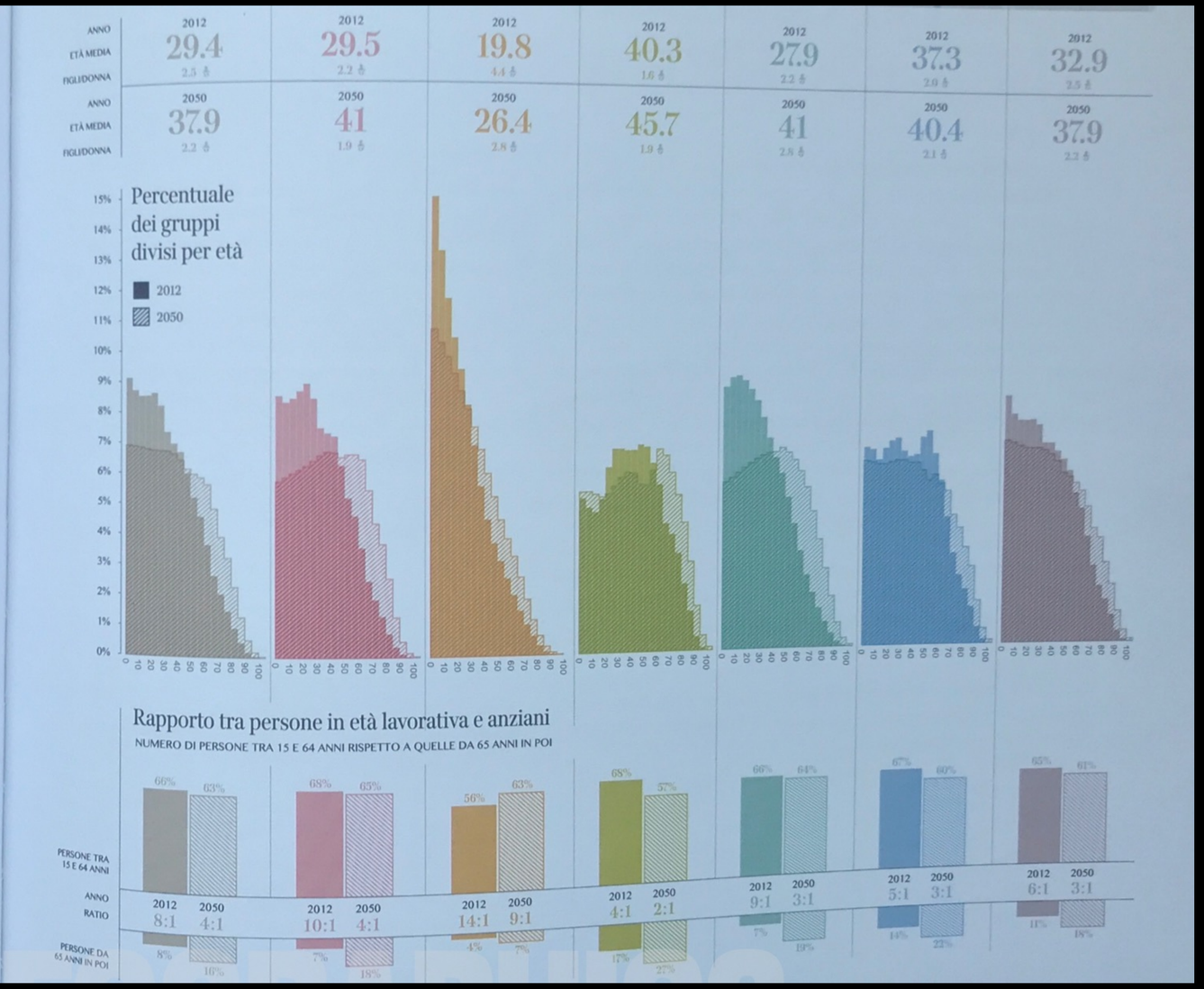
EDITED G.COLIN, A.TROIANO

RIZZOLI, CORRIERE DELLA SERA

2014

INFOGRAPHICS

La visualizzazione dati è a cura dello studio Hyperakt di New York, specializzato in progetti di design sociale e guidato da Deroy Peraza e Julia Zeltser. Il grafico è stato pubblicato su «la Lettura» il 20 gennaio 2013.



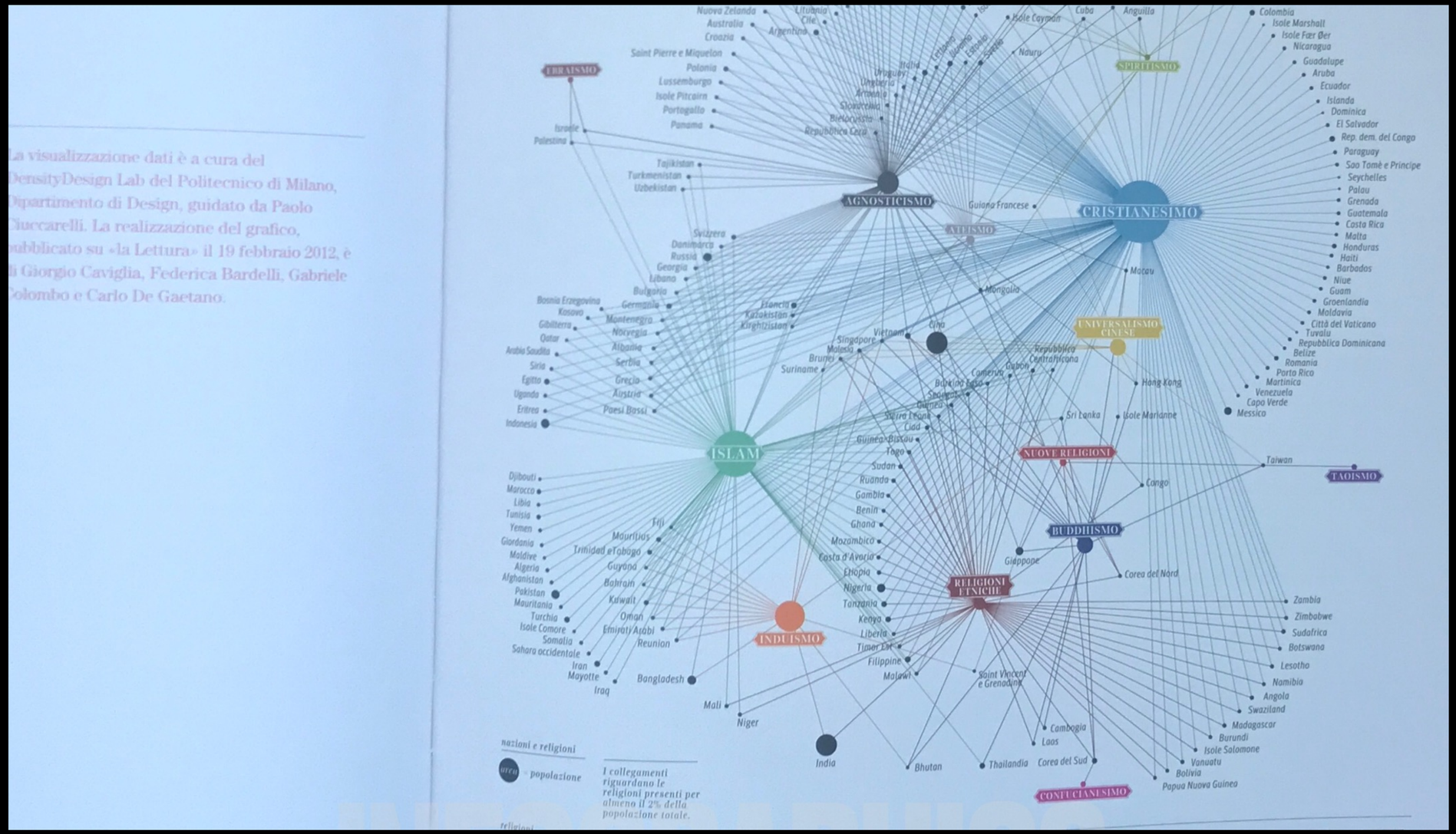
LE MAPPPE DEL SAPERE.VISUAL DATA DI ARTI, NUOVI LINGUAGGI, DIRITTI.L-INFOGRAFICA RIDISEGNA LE CONOSCENZE

EDITED G.COLIN, A.TROIANO

RIZZOLI, CORRIERE DELLA SERA

2014

INFOGRAPHICS



INFOGRAPHICS

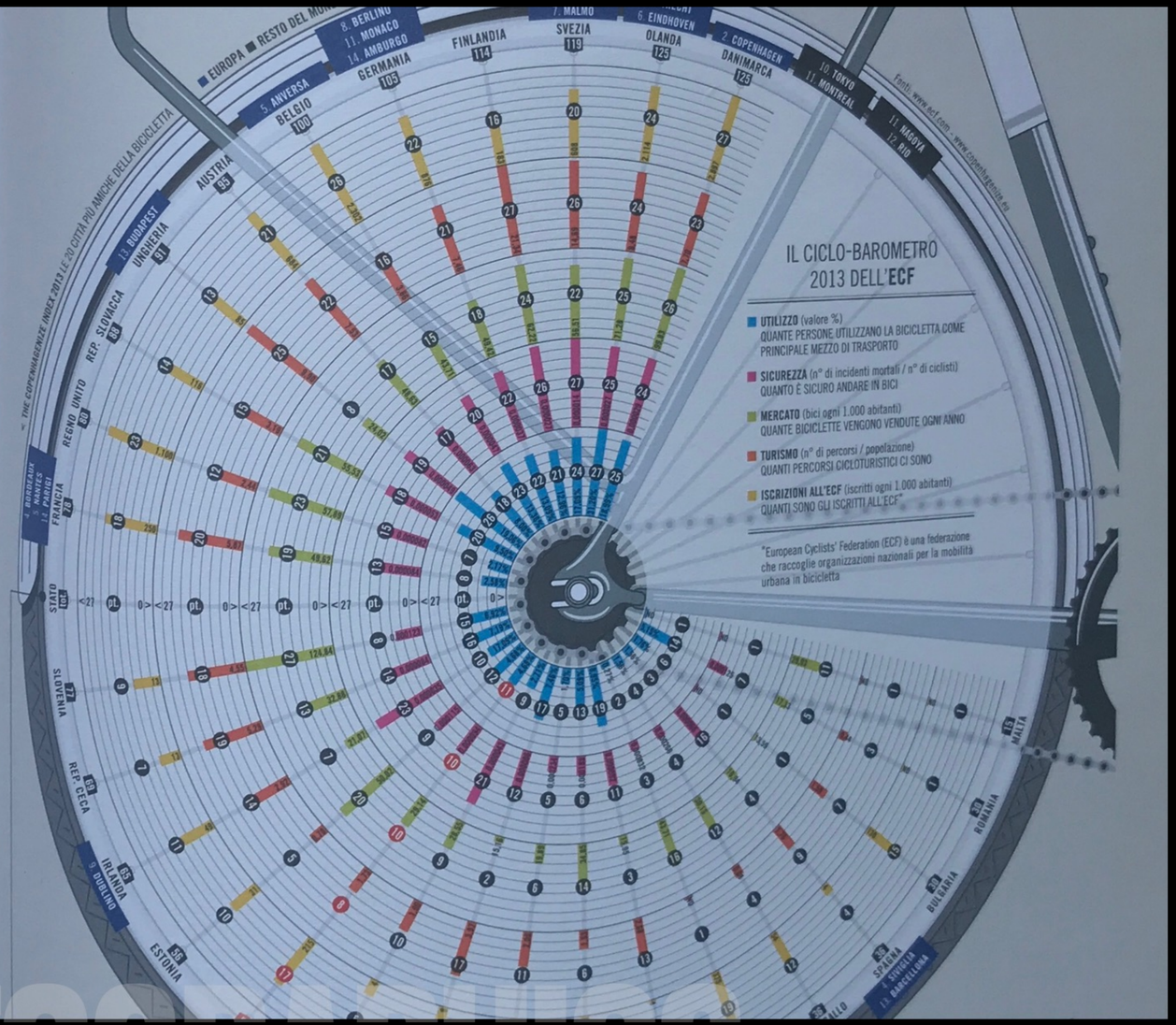
LE MAPPPE DEL SAPERE.VISUAL DATA DI ARTI, NUOVI LINGUAGGI, DIRITTI.L-INFOGRAFICA RIDISEGNA LE CONOSCENZE

EDITED G.COLIN, A.TROIANO

RIZZOLI, CORRIERE DELLA SERA

2014

La visualizzazione dati è a cura dello studio bruno di Venezia, diretto da Andrea Codolo e Giacomo Covacich, specializzati in design dell'informazione. Il grafico è stato pubblicato su «la Lettura» il 9 febbraio 2014.



LE MAPPPE DEL SAPERE.VISUAL DATA DI ARTI, NUOVI LINGUAGGI, DIRITTI.L-INFOGRAFICA RIDISEGNA LE CONOSCENZE

EDITED G.COLIN, A.TROIANO

RIZZOLI, CORRIERE DELLA SERA

2014

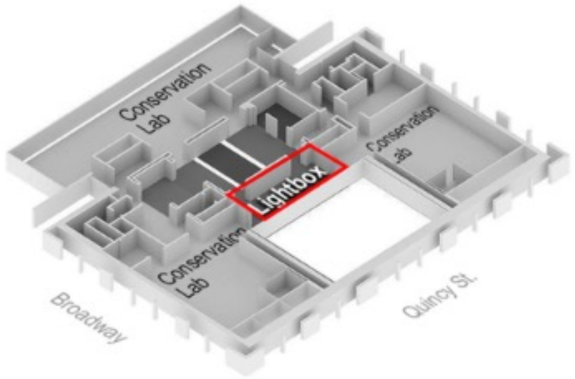
INFOGRAPHICS

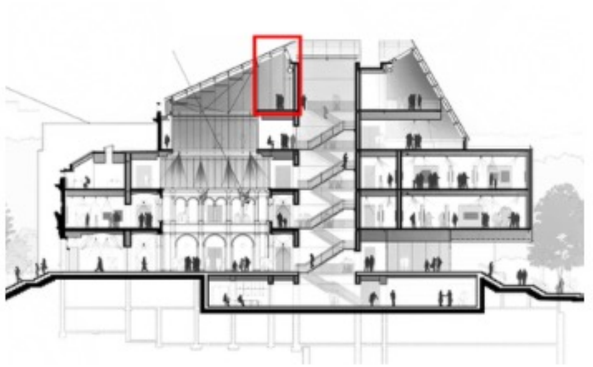
DH AND DESIGN

Jeffrey Schnapp

HARVARD EDEX

ON GOING





Lightbox Gallery: Introduction

http://apps.harvardartmuseums.org/lightbox/index_info.html

The Lightbox Gallery is a venue for digital experimentation—a space for projects that respond to the museums' collections through new media and emerging technologies. Developed in collaboration with faculty, staff, students, and visiting artists, the projects showcased here use digital tools to reveal connections between objects and to play with traditional forms of display. Some of these projects are responsive, allowing users to navigate and manipulate the collections; others are cinematic, transforming the museums into a landscape of digital performance. Exploration in the Lightbox is cross-disciplinary, merging a gallery experience with a digital lab.

LIGHT BOX GALLERY HARVARD MUSEUM

BIBLIOGRAPHY

- Burdick A., Drucker J., Lunefeld P. Presner T., Shnapp J., *Digital_Humanities*, MIT Press, Cambridge (MA), 2012
- Capezzuto S., Il design della conoscenza. Intervista a Jeffrey Schnapp, *Il Lavoro culturale.org*, 6 ottobre 2017, (https://www.lavoroculturale.org/intervista-a-jeffrey-schnapp/stefano-capezzuto/2017/)
- Gardiner E., Musto, G., *The Digital Humanities: A Primer for Students and Scholars, Cambridge Univeristy Press,* Cambridge (UK), 2015 (https://amzn.eu/7k8qe73)
- Harvard edex course: Introduction to Digital Methods for the Humanities, free Online Courses by Harvard, M.I.T. https://www.edx.org

WEBGRAPHY

- http://apps.harvardartmuseums.org/lightbox/index.html
- http://www.verbovisualevirtuale.org

Informatica per le Digital Humanities
Lecture 2
February 24 2023

Texts

Slide 2

If you remember, my aim was tripartite.
Firstly, I wanted to define the term "Digital Humanities," starting from questions and doubts surrounding the term, and to show that there are several divergent points of view.
We dealt with that point in the previous lecture. Now we have to move on to the next two points.
We will examine the chronological development of the discipline, with an overview on the most significant projects from the beginnings until now. We will try to build a timeline of Digital Humanities.
Then, we will discuss some methodological questions concerning the relationships between two worlds: Computer Science on the one side, and Humanities on the other. My purpose will be to show you that there are both encounters and collisions.

Slide 3

The earliest examples of what we now call digital humanities were placed under the name of "computational humanities", "computing and the humanities", "humanities informatics". And so in the course of many decades, as we have said before, the terms evolve, but at the very core of these efforts, until now and foreseeably in the future, there is a conversation between computational techniques, technologies and platforms on the one side, and the kinds of questions that characterize the humanities disciplines on the other.
We can talk of a DH History because there is an official engagement between the Humanities and Computational Science starting from the late 1940s with Father Busa.

From the 1940s until the 1980s we can observe several scholars focused on opening archives and databases to a wider audience. In particular, librarians and information science specialists worked on humanistic corpora and developed file formats and machine-readable records for this purpose.
Let us now look at a few significant examples that need to be inserted in a precise context of technological and cultural development in order to be better understood (something I will leave to you). Here, we will focus on what we can profitably use in the next lessons about texts.

Slide 4

Roberto Busa, a Jesuit father, is considered the pioneer of Computational Linguistics, a discipline focused on the development of formalisms describing natural languages and functions, and their subsequent transformation into executable programs on a computer.

In 1946, while writing his master's thesis, he matured the idea of a precise and complete verification of the lexicon of Thomas Aquinas. It was the beginning of a life-long research, which took off from a very specific starting point: in Saint Thomas' writings, the concept of "interiority" is present in the form "being in"; however, an automatic search for the occurrences of the particle "in" was not available at the time.
In 1949, during a trip to New York, Busa had the chance to present his idea to Thomas Watson Sr., founder of the IBM Corporation, who decided to support his project. The goal was to perform a full verification of the lexicon of St. Thomas Aquinas in order to obtain an authentic interpretation of his thought based on the results of such an analysis, finally cleansed of the innumerable encrustations due to centuries of comments and interpretations.
The use of IBM technology was, at first, punch cards and then magnetic tapes designed for the classification of words.

Slide 5

In 1980, after thirty years of work, the printed edition of 56 encyclopaedic volumes of the the "Index Thomisticus" saw the light of day, an imposing work which gathers the entire production of St. Thomas Aquinas in a format readable and manageable by computers using the methodology developed by Father Busa.
This first project has three decades of development behind it. This depends on its experimental and pioneering character, but the long-lasting effort is also constitutive of this kind of works, as we will see later.

Slide 6

In addition to the imposing work of Father Busa, the first statistical processing of the Divine Comedy was developed in the mid-1960s with the help of the 14018 and 7090 processors, and the first Electronic Dictionary of the Italian Language (VELI) was completed in 1989.
The Dante Project was focused on taking the six centuries of commentaries on Dante's Divine Comedy and transforming them into a database that would allow scholars to ask questions on the interpretive history of the poem that are very difficult to ask when working with non-digital editions of commentaries.

Slide 7

As we move from the 1980s to the present, what we see is a growing portfolio of practices, a growing set of experimental practices that characterise the convergence, but also the collision between humanities disciplines and computational techniques and tools.
So, "digital humanities" becomes the defining label for this field of experimental practice, especially with the rise of two key factors.
Firstly, the emergence of the Internet as the ultimate public space of our era.

Secondly, the increasing importance of personal computation and personal smart devices as they become not just simply desktop equipment in offices, but fundamental features of our everyday life.
These two factors are at the basis of many projects of enormous digital libraries, such as the forerunner Project Gutenberg or the now mainstream Google Books.
Announced officially in 2010 (but secretly launched on 2002), today the Google Books Project can really invite users to "Search the world's most comprehensive index of full-text books", as advertised on its homepage. Sensitive copyright issues notwithstanding.
In August 2010, Google announced that by the end of the decade it would scan all known existing books: all 129,864,880 of them for a total of more than 4 billion digital pages, which translates to 2 trillion words.

Slide 8

We can affirm that one of the most interesting phenomena to emerge from the intersection of so different disciplines in the DH is the definition (or at least an attempt of a definition) of fundamental methodological elements in digital scholarship.
Digital scholarship is an umbrella term that encompasses visual humanities, digital history, new media, and a variety of computational methods in general.
Digital scholarship, by its very nature, places greater emphasis on interdisciplinary collaboration than traditional approaches. For instance, traditional approaches to scholarship often emphasise the skills and research of an individual scholar. In contrast, due to the inherent complexity of digital tools, digital scholarship often makes it necessary for researchers to work in collaborative, multidisciplinary teams.

Slides 9 and 10

"To do something really interdisciplinary, it is not enough to choose a 'subject' and gather two or three disciplines around it. Interdisciplinarity consists of creating a new entity that does not belong to any discipline." (Roland Barthes)

Slide 11

Last week we tried to understand the contributions that one discipline can give to the other.
To insert a preposition "with", and to propose a term like "DIGITAL with HUMANITIES" means to shed light on a real interchange between two entities.
An interchange with a number of contradictions, which we need to be able to tackle and, if not solve, at least understand.
We have focused on the necessity of a discipline that is concerned about the interaction of human beings and their unpredictable behaviors, not easily MEASURABLE nor PREDICTABLE, with a discipline (digital technology), whose statements and endeavours lie within the realm of MEASURE and REPETITION.
So, if the humanities are to give digital information technology a necessary (or inevitable) dose of uncertainty, what can digital humanities give to humanities? Here some some questions.

Slide 12

If you continue reading in the textbook, you will find this:
“While some believe that the digital is fundamentally changing the work of the humanist, others continue to believe that the digital merely helps humanists to work better. Some even believe that the digital may be undermining the fundamental nature of this work. Many humanists tend to view the digital humanities as a methodology that brings the tools and power of computing to bear on the traditional work of the humanities.
Computer scientists tend to view the digital humanities as the study of how electronic form affects the disciplines in which it is used and what these disciplines have to contribute to our knowledge of computing.”

This is a chance for us to reflect on what was brought by the interaction between different worlds:

- it has created in the humanities the possibility of opening up to a wider audience without losing the complexity and accidental nature of their ways;
- it has helped scholars manage such complexity in their research efforts;
- it has brought visual learning to the fore (i.e. DATA VISUALIZATION techniques) with a very interesting graphic contribution coming from the development of digital disciplines and, in this way, it has added creativity to the research.

Slide 13

This is a collection of contemporary art, seen as a whole and in its smallest details, out of and far away from the places where it is stored.

Slide 14

Now some images related to Data Visualization.
We are talking about a series of infographics appeared in an insert of the Corriere della Sera newspaper.
This may be an oversimplification, but there is a certain truth in saying that the image has the gift of synthesis and the word that of analysis. So, viewing huge amounts of information in one go is no small feat.
Here you can see different styles of infographics. The data can be the most disparate, of course. From the spread of rap music, to the writers of mystery books, or to the recurrences in Harry Potter, to the disappearance of the book heritage, to the evergreen classification of the Beatles, to demographic statistics, to religions in the world, and so on.
As you can see, often in the visualization signifier and meaning can coincide.

Slide 22

Once again, Jeffrey Schnapp

VIDEO TRANSCRIPTION
JEFFREY SCHNAPP: Digital humanities projects
over the course of their long history from computing in the humanities,
computational humanities, humanities informatics,
all the way to contemporary digital humanities,
have tended to divide into two main categories.
One is infrastructure building, or at least that's how I like to call it.
And the second is creative, expressive, critical, and experimental practice,
sometimes at the boundary line between art and scholarship.
The first category really characterize much
of the work that happened between the era of Father
Busa and the internet revolution.
It involved the construction of large-scale resources
and supports for humanities scholarship.
That's certainly the case of the Index Thomisticus.
It's the case of the Dartmouth Dante Project.
It's the case of NINES and The Rossetti Archive.
They're all projects that are striving to create
resources that build on existing traditions of humanities scholarship
and practice.
The second category, a category that's more
focused on expressive, experimental, critical forms of scholarship,
really takes that infrastructure building in a new direction.
These are attempts-- I would describe them as attempts--
to create new genres of scholarship, new cultural forms, new models
of communication that leverage the power of digital media
and computational techniques to tell new kinds of stories,
to expand the compass of scholarship, to reimagine the boundary
lines between the humanities set of disciplines
and other disciplinary clusters, like the social sciences
and the natural sciences, but also to reimagine
the relationship between scholarship and forms of public discourse,
cultural practice and experimentation, forms of argument,
and the civic and public sphere.
So in short, I think these two broad domains of digital humanities practice,
even as they overlap and build on one another,
have two fundamentally different focal points.
The one is really supporting and sustaining
traditional forms of inquiry.
The other is inventing new forms.
So one of things that really excites me about the work that's happening today
in I think at the frontiers of the digital humanities
is work that is really thinking in a whole new way about scale,
about the scale of argument, about the scale of human experience,
about the scale on which culture unfolds.
And in a number of projects, among our projects of my own, of my lab,
the metaLAB here at Harvard, some of them

are projects being carried out by people like Lev Manovich, who is really
one of the pioneers in the digital media and sort
of computational culture field, so-called cultural analytics.
These are all projects that do leverage the power of cultural corpora of data
sets to tell stories that aren't the kinds of stories
that have characterized the history of civilization up until our time.
Those are stories that occur on the sort of scale of human experience.
We as humans are very brilliant, agile, at keeping
in our field of vision, in our frame of mind,
in our visual field, maybe five, six, 10,
a hundred actors at a single moment in a single frame.
But once we start to scale beyond that human scale,
we have a hard time processing that kind of information.
But data and databases are indifferent to that question of scale.
They work just as well at the scale of millions of objects
as they do at the scale of one, a single object.
And inventing storytelling modes, finding ways
to create experiences that are meaningful experiences of very, very
large fields of objects, large cultural corpora,
is a domain of experimentation that has characterized
a lot of the really cutting edge work over the past decade or two.
I'll mention two projects that I think are pretty interesting in this regard.
The first is a project of Lev Manovich's that has the title Selfie City.
And what it sought to do was to capture the collective production of selfies
throughout the world and to use the geotagging and the time tags
on each of the photographs to tell a story about the places
where people produce these vast seas of self-representations
through their cell phones, to aggregate them, and to represent them.
It's a project that you could describe as an art project.
That's certainly how it expresses itself through these vast kind
of constellations or clusters of selfies produced
in different physical locations over the course of a day.
But you could also describe it as a project of urban mapping
because what it does is maps the demographics of the cities.
What did people represent?
How do the selfies vary from one location
to another, from one moment of the day to another, from breakfast time
to lunchtime to after work?
So it's also a demographic study if you like, a kind of live,
living sociology of different cities.
What do the backgrounds tell us about the kinds of cultural context
or the sorts of visual values that characterize
each of these places and times?
Those are the kinds of questions that hover around a project like this.
It doesn't provide answers to all those questions.
But rather what it does is leverage the tremendous power of these resources
to tell a very big story, a story that's zoomed out to the scale of millions

of people and multiple sites.
The second example I'd like to mention is the project
that the metaLAB designed for the Harvard Art Museums
when the museum reopened a few years back.
And it's called the Lightbox Gallery.
And the idea of the Lightbox Gallery was to create a place, a kind of immersive,
interactive gallery, where the database that visitors to the museum experience
online where they can see the collections,
but they, in particular, can explore those collections interactively,
to turn that into a physical embodied experience
at the conclusion of a visitor's visit to the museum.
And what the Lightbox Gallery does is it consists
in a wall of monitors on one side and a projection
system on the backside behind the visitor so to speak.
And the visitor interacts with the monitor wall through an air mouse.
And on the monitor wall, which is facing you
as you walk into the gallery having completed
your tour of all of the floors of the museum,
you see not a single art object.
You see all 1,803 objects that you have already visited face-to-face
in your tour through the galleries, but all
in the form of a kind of mosaic of thumbnails.
And as you point the air mouse to any of those objects,
that object is pulled up.
And what you see is the data record.
You see it as a digital object.
In other words, it's not a question of simply replicating the experience
that you already had face-to-face and creating
a kind of impoverished version of that experience on a screen,
but rather showing that that object as it exists
in a database is a different object.
And if you move around the air mouse to any
of the fields that make up the data record,
the projection system on the back wall actually
shows where that object stands based on that particular field
in the whole of the 1,803 objects that are on exhibit inside the museum.
The point of this is to show what data does
best, which is unlike the richness of that face-to-face experience
of a cultural object, what a data set does
is really to open up not that kind of qualitative experience that
is such a rich sensory experience of an art object,
but rather to open up a multitude of perspectives,
to look at the same object from any number of perspectives, all
of the fields that make up a database are available to you as points of entry
and then to see relationships, to see patterns.
So if you click on Provenance, like where did this object come from?
Or you click on how many times has it been viewed online?

Or you click on when was it produced?
Where was it produced?
What material is it made of?
What you see is the object as a set of relationships.
Sometimes, I like to describe the Lightbox Gallery
as a kind of social network of objects.
So rather than seeing an art object as an individual object, suddenly
you see that it has a family, and it has friends,
and that in every field that you look at, it has a position in relationship
to the other objects.
So it's those networks, those relationships,
that become the object of an experience.
And that's an experience that's not competing
with the face-to-face experience of an artwork,
but rather adding a kind of depth, a kind of value, that provides,
in the aggregate, it provides a portrait of the institution
as a whole, a collective portrait, a kind of portrait that incorporates
a view that, in a sense, is not even available to the curators
and the museum administrators who have access to all the back rooms
and so forth.
It's that kind of institutional portraiture that I think
is really an exciting dimension of the digital age
that we can design experiences of history or storytelling
or an institution that has that-- that have those kinds of components of scale
that allow us to zoom from the macro view
all the way down to the single object and then back again effortlessly.

Slide 23

If you prefer to have a look, here is the site.
Here on the slide you can find the correct link of the project quoted by Schnapp.
But here, online, you can’t find anything special, except for a definition concerning this new interesting age of studies based on RELATION, PROCESS, and CROSS DISCIPLINARY, something that is going towards a new discipline’s definition.
From DIGITAL HUMANITIES to DESIGN OF KNOWLEDGE?
The question, as usual, is open.