#DHCH Digital Humanities Summer School in Bern

------Wednesday 26 June 2013

Welcome by : Natale (@infoclio), Clivaz (@Cclivaz), Stolz, Zala (@SachaZala), G&I (@gischweiz), Kaplan (@frederickaplan)

==> You can write in any language. Please participate, this will be useful only if it is a collective effort. <==

Ray Siemens(@RayS6): Understanding and Responding to (Inter)Disciplinary Change

"Do you mind if I'm informal ?" ...Bearded guy from Canada ;)

1. Humanities 2.0
has some resonance in Nothern america, cf papier by Davidson : promise, perils, predictions of humanities 2.0, PMLA, vol. 23/3: http://fredgibbs.net/courses/digital-history/readings/Davidson-Humanities2.pdf
Series of article of the New York Times under the label "Humanities 2.0" : http://topics.nytimes.com/top/features/books/series/humanities_20/index.html
Big tent version of DH...

2. Envisioning the Digital humanities
R. Siemens presents a general overview of DH fields of application (humanities disciplines, various digital techniques, etc.) and international organisations (ADHO, NEDIMAH, ALLC, etc)
Difficulty to draft an all-encompassing definition
https://twitter.com/GrandjeanMartin/status/349865640171020290
Ref: S. Schreibman, R. Siemens, j. Unsworth, A companion to Digital Humanities, Blackwell, 2004. Online: http://www.digitalhumanities.org/companion/
Ref: S. Schreibman, R. Siemens, A companion to Digital Literary Studies, Blackwell, 2007. Online: http://www.digitalhumanities.org/companionDLS/

3. Defining & determining actionable model
Manifesto style thinking is difficult to implement... But is there a need for a specific disciplinarity?
https://twitter.com/tla/status/349865219222290432
https://twitter.com/pauldoshea/status/349867141501157376
https://twitter.com/frederickaplan/status/349866970662973440

Steps that could be taken :
Mapping the methods : Mapping the methodological commons onto extent structures around us => to bridge what's important (https://twitter.com/xaentenza/status/349868167453085700)

4. Engaging key trends
Future is hard to predicts, but we can look at the directings it seems to be going
- Content modeling of data (representation)

- Process modeling with tools (analysis)
- Communication / dissemination
- Access to large data - in usable formats
- Crowdsourcing -> [comment]: Crowdsourcing is a largely myth. Successfull crowdsourcing projects like Wikipedia or Transcribe Bentham rely on a very limited number of very productive and dedicated users. Ask a million to

- More softwares, and easier to use

Willard Mc Carty, The "methodological commons":
http://rsta.royalsocietypublishing.org/content/368/1925/3779/F1.large.jpg

5. Responses ?
Institutionnaly, administrative and community changes as consequences of DH
This could result in some reluctances against DH because it implies all these changes
Example of good practices:
etcl.uvic.ca: http://etcl.uvic.ca/
dhsi.org: http://www.dhsi.org/
The future of the book of the perspective of its past: inke.ca: http://inke.ca/

6. Discussion
Discussion about the wordcloud Prof. Siemens showed in his lecture (wordle done on a MIT conference's abstract book texts, byt the organizers). Function words are gone, are kept only content words.
Top 50 is visualized.
Why are Library, Archive, and Museum absent ? Then the discussion switches to what is the algoritm beyond it ?
Some possible biais :
- you can talk about something without using the word.
- ?

Then what is the usefullness of such wordcould if the algorithm beyond it remains hidden ?
Except being good looking. Digital is very big but what does it tell us ? Not much!
Remarque intéressante de Claire Lemercier : "Not as trivial as expected," ne pas se focaliser sur les éléments qui apparaissent évidemment au premier plan, mais s'intéresser aux marges.

Really interesting discussion, "very productive & engaging", "can only happen with digital humanists", says Prof Siemens \o/
https://twitter.com/squintar/status/349874587045072897
Outil alternatif conseillé sur Twitter : http://www.iramuteq.org/

Question from E: Pierrazzo: Is "Big Tent DH" a good or bad thing ? Are we dilluting our specificities ?
Siemens: Did DH stops where humanities stops ? or is the "big tent" bigger ?
Is the Merging of Social Sciences and the Humantities inevitable in the digital age (beyond the age of the book) ?

[Course] Susan Schreibman (@schreib100), History and future of DH

Questions without answers:
- Is DH a field?
- WHo is a digital humanist?
- What is DH ?
- We speak about "digital humanities", but in Switzerland "Humanities" is not an academic category...<= this is a real question ! I don't think it is in France either :-/

[Participants are invited to give their 1 line "elevator-pitch" that best defines DH]
- "Dust meets geeks" by Andreas Perret
- "Humanities using digital tools"
- "Putting Shakespeare on iPads" : Laura Mandell cited by Schreibman. Discussion here : https://twitter.com/pauldoshea/status/349889786133221376

Re-thinking the object life-cycle
Recalls Roberto Busa as founder of DH discipline: http://en.wikipedia.org/wiki/Roberto_Busa
Busa-Waston cooperation to establish a concordance of St. Thomas Aquinus with punched card IBM machines. Cf. http://www.guardian.co.uk/higher-education-network/blog/2011/aug/12/father-roberto-busa-academic-impact

So, Digital Humanities: a field, a tool, a discipline? All of these? None? "I do not think these words mean what I think they mean."
https://twitter.com/twinkling_sheep/status/349891702976290819

Google Ngram montre émergence/décollage de Digital humanities en 2000, remplaçant Humanities computing qu'on voit à partir de 1967-68

[useful_link] Where you can ask any question DH Questions and Answers:
http://digitalhumanities.org/answers/

"Binary code, truly universal language".
"Moving from reading and critiquing to building and making" : Digital Humanities, it's about "making" ! https://twitter.com/GrandjeanMartin/status/349893771334393856
Developing new tools and softwares...

To be read : Bibliography and the Sociology of Texts by D. F. McKenzie :
http://www.amazon.com/Bibliography-Sociology-Texts-D-McKenzie/dp/052164495X

Débordement des rôles - Digital Humanists <->Editeurs
RefHow should DH work be evaluated: www.mlajournals.org/toc/prof/2011/1

.DH as pedagogical practice :

Area: Text visualization and Distant Reading:
Biggest corpus yet: Google Books Project
Example of visualiation of language uses at the macro level with Google NGrams: http://books.google.com/ngra
Other example of visualisation of Wikileaks Iraki Wars Data. --> Is Wikileaks a DIgital Humanities Project ?
More example of Distant Reading Projects about Agatha Christie and an "obscur Irish Poet"

Area: Virtual Worlds
Exploring virtual worlds and reconstructing historical settings in order to answer historical research questions!

Ref: Being and Seeming: the Technology of Representation by Richard Powers. Online: http://www.wildethics.org/essays/being-and-seeming.html

[useful_tool] http://www.textarc.org/ = A TextArc is a visual represention of a text—the entire text (twice!) on a single page. A funny combination of an index, concordance, and summary; it uses the viewer's eye to help uncover meaning.
[ref] "Raw Data" Is an Oxymoron (Infrastructures) by Lisa Gitelman (http://www.amazon.com/dp/0262518287)

-----Thursday 27 June 2013

[Course] Elena Pierazzo(@epierazzo), Digital Textual Editing

see blog : http://epierazzo.blogspot.ch/2011/10/role-of-technology-in-scholarly-editing.html

Editing is an old practice (Library of Alexandria) : methods to make **correct** a message.
Message goes from a sender to a receiver, by channels. Loss of message, noise (There can be noise in the writing system, in the writing conventions, in the style of writing, in the support, in the screen colors, in the pronunciation...)
But what is **correct**, defined by the role assigned to the physical support of the text.
Document = physical object => picture of a document is only a representation
Text = more difficult to define. A linguistic architecture with a meaning potentially understandable by some receivers that have the caability to decipher the code. Can be immaterial.This alllows to be transmitted. But they are also embodied in physical objects, not/difficultly transmitted.

What is digital editing ? New method or new discipline ? Lost of things have changed, the way we read, edit, publish...
EDITING

- 1st Example : a project at KCL, edition of inscription, like puzzle of stones

Projet de numérisation des inscriptions de la Rome tripolitaine: http://irt.kcl.ac.uk/irt2009/ (intéressant renversement de situation: des inscriptions sur pierre sont peut-etre sauvées des destructions de la guerre en Lybie par la numérisation (alors que la pierre est un média plus pérenne, à première vue).
We don't know how long a digital edition will last, should we print them out ?

- 2nd example : fac simile : you wil never see the transcription in its own without the fac simile, it is only functionnal, to help you read easily the text. Ultra-diplomatic edition. (http://en.wikipedia.org/wiki/Diplomatics)

- use of Jane Austen's manuscripts as another example of the types of projects that incorporate digital editions as a tool towards scholarly editions:

- http://www.janeausten.ac.uk/index.html]

-

READING
Do we read the digital editions on screens ? Not for long hours.
Not for all : Les ebooks sont en nette augmentation aux USA, progression en Europe, mais très largement absents d'Europe de l'Est et d'Afrique http://paidcontent.org/2012/06/12/what-will-the-global-e-book-market-look-like-by-2016/

Scholarly editions : Are scholarly editions meant to be read ot to be used as Intellectual tools/explored? Do we really "read" them ?
Codex Sinaiticus: http://codexsinaiticus.org/en/manuscript.aspx
van Gogh http://vangoghletters.org/vg/
What do the readers want ? The answer is : Printed books ;) (Cf : Porter 2013 survey with mediavelists: http://www.scholarlyediting.org/2013/essays/essay.porter.html
Are scholarly editions to be read or used? Example of TSE Waste Land App (where it is "literally" being read) and also the Shakespeare sonnets app

Preparing the editions : 2 main approches
- (a) evolution of the editor
- (b) the computer as mechanic research assistant : magic/black boxes

Editor as encoder/programmer/web designer lives in the believes
- that enconding is interpretation,
- encodng make explicit our understanding,
- encoding is a way to represent research

(a) Advantags of these approach (ex TEI)
- keeping the annotations and the documentations of the editorial work exactly linked to the word it belongs
- Document at every level, all the details of the work
- maximum flexibility in the output
- possibility of running statitstical queries
- quality control
- international community of users
- open to innovation
BUT : you have to learn so many things ! what every editor should know is a very long list.

XML + TEI + XSLT + HTML + CSS... (but "HTML and CSS are so 2005"...) + web design + databases +...

then let"s look at the magic boxes :
(b) Computers do all for you. Kinda. But you don't understand what you're doing. Many different traditions, only TEI is adaptable to all.

What solution, then ? TEI tools are rigid, and yet, TEI is also deemed "extremely flexible"
Wake up ! Social edition ! You are not alone ;), wisdom of the crowd/crowdsourcing --> ha ha, I didn't realize you were typing on the lecture: I thought you saw me sleeping... :S => #MOUAHAHAH hu hu

Transcribe Jeremy Bentham: http://blogs.ucl.ac.uk/transcribe-bentham/
http://www.transcribe-bentham.da.ulcc.ac.uk/td/Transcribe_Bentham
"The Devonshire manuscript/Farewell all my wellfare":
http://en.wikibooks.org/wiki/The_Devonshire_Manuscript
Wikibooks, an example of the wisdom of the crowd: http://en.wikibooks.org/wiki/Main_Page

Then if everybody is doing it, what the editor is supposed to do ? Change job ? --> "crowd sourcing" (the job of the poor research assistant for the overlord editor)
- Setting up the Infrastructure
- Selecting/organizing the material
- Checking the quality of the transcription
- Form transcription to edition

Documentary editions and computers -- computers have greater flexibility that allows for changes and more capabilities in editing: http://codexsinaiticus.org/en/
Digital editions are getting more imaginative and interactive : from digitised to digital editions

"From Theory to Encoding"
OCHO model: the objects are chapters, sections etc. But not pages or any other material references which are not useful. Little to do with real texts.
OCHO Model not interested in the material dimensions of texts. Cannot have linguistic, semantic, or literary dimensions simultaneously
https://twitter.com/pauldoshea/status/350161586452828163

Editing is about texts, but sometimes the process to make the document is at least as important as the text that it contains. The text can be determined by the document.

- 7 different reconstructions of Beowulf's manuscript "gatherings"
manuscripts are messy, and DH is a way of reassembling manuscript and organizing archive

What do we do with non linear texts ? Picasso, Apollinaire...medieval texts (5 columns = 5 texts in one) ?
Editing is all about texts, but documents are at least as important as the texts they contain.
Document-type dependent? https://twitter.com/pauldoshea/status/350162585456361475
=> Document based encoding, support for messy manuscritps, helps for genetic editing/editions

Non-linear example (picture):
https://twitter.com/GrandjeanMartin/status/350164196094255104

Example : Madame Bovary http://www.bovary.fr/ = looks to me like an "ultra diplomatic edition", so 2000 ;) => new modules
The new module has 3 parts :
- document encoding
- transcription enhancement
- ?
each manuscript can be given "exact spatial coordinates" in the transcription
Does it work ? Well, we'll see ;)

Prototype : a notebook of Proust: research.cch.kcl.ac.uk/proust_prototype
Examining the writing and reading sequences (different)
Main research question : How to make understandable editorial work for a larger public ?
How to present the genesis ? Digital edition that WOULDN'T be the transposition of print editions on the screeen !
"what can a computer do that a book cannot do?" The goal is to achieve what cannot be achieved via print. https://twitter.com/pauldoshea/status/350165312177573889

My inspiring principles :
- See Gabler http://compositiongenetics.org/navigate.php?page=base --> demonstration of an interactive "manuscript transcription tool" that magnifies sections of Ulysses manuscripts and inserts Gabler's transcriptions
- Computer games : if kids can play with evolution, why wouldn't they play with edition ? Gamification !

"Hyper" Nietzsche collective "manuscripting": http://www.nietzschesource.org
@radusuciu:10:24 yes. the animated one. @frederickaplan found this screenshot:http://t.co/PVElyvsBEN --> for the exact pg being discussed http://www.compositiongenetics.org/bksailehwgabler-33

Proust Cahier 46: http://t.co/2SE7Q612xJ - allows readers to see the text /(or document?) actually changing in time and space.

Digital textual ed. changes division of labour, the way we think of our work, the way we think of ourselves, and the reasons how/why we do our work.
Conclusion : Digital editing: new medium ? new discipline ? ---> "for me, not clear yet" !

[Course] Ray Siemens(@RayS6), Social Knowledge Construction and Creation in Literary Studies Environments

Social Knowledge Construction and Creation (including Notes from Ray Siemens copied from http://goo.gl/LT8mR)

- academic engagement around social media

Framing questions:
1) what is knowledge?
What I've learnt. Can intuitive knowledge be shared?

2) What does 'social' mean, in a computational context?
Some of the answers :One to one, one to many, many to many communication. p2p

3) what is the first word that comes to mind when you hear the name "wikipedia"?
Some of the answers :Collaboration, Encyclopaedia, Halleluja, totalitarism, Inaccurate, underappreciated, lack of authority (speaker)

Context
Mapping from McCarty, again http://www.allc.org/node/189
Web 2.0 practices and standards have encouraged scholars to rethink from a user perspective the design of the scholarly edition—the fundamental textual form in literary-oriented research (Vetch; Robinson). Books have always been inherently social media as Jerome McGann argues, suggesting their suitability for a form of social editing (McGann, The Textual Condition) in a digital edition able to reflect through careful design the dynamic relations inherent to textual production and reception ("From Text to Work: Digital Tools and the Emergence of the Social Text"; see also Liu). Community and collaboration are integral to scholarly knowledge creation (Fitzpatrick) and Siemens *et al.* integrate within the social scholarly edition collaborative electronic tools for annotation, user-derived content, folksonomy tagging, community bibliography, and text analysis. The modelling of the scholarly edition through applied social knowledge creation tools (ASKCTs) answers the call for polyphonic interpretation of multiple readers within an edition (see Smith), wherein the editor become a facilitator of knowledge creation rather than a didactic authority . . . .

Trends:
- increased data, accelerated communication, better ways of asking question
+ workflow speeding (basic and advanced)

Converging to public investigation, creation of tools which may change the questions we ask.

Goals of course: seek understanding of trends (past -> present -> future), expand the community to build expertise. Collaboration as best approach. (working with others to increase understanding)
Learning through doing/making.

How We Approach Knowledge Construction, Intellectually
- Knowledge has always been messy !
- How knowledge is conveyed and how authority is conveyed in traditional academic circles.
- Knowledge production is age-old; what's changing is our way of interacting in this community. Very different in different areas, no clear sense of what a book is, knowledge is.
- Print culture has never been something monolithic and singular
- Knowledge seems to be build up over time by means such as conversation manuscript circulation, example but nothing is defined: multiplicity
- Remember internet has been messy ! Some people have tried to make it authoritarive, the same message from one to many. Didn't work. Became even more messy with google docs...(enfin il parle des google doc modifiables, hein)

- Telephone game (en français, called "téléphone arabe", http://fr.wikipedia.org/wiki/T%C3%A9l%C3%A9phone_arabe). Corresponds +/- how social media works.
Canadian telephone for @RayS6 : what word shall we use?
but also Chinese Whispers !! RE
https://twitter.com/twinkling_sheep/status/350184261703315456
--> demonstrates the perils and wonders of transmission (the game of Telephone, that is)

Humanists have always engaged this sort of thing : (internet...)
- embracing (cp. Haraway; Aarseth),
- studying (Liu),
- problematising (effects of social media on scholarly practice: Mrva-Montoya; challenging power of searchmachines and how the unify knowledge: Introna and Nissenbaum; Berry),
- acting ?

In web 2.0, we have to understand our role is changing. *[==> (Who is our/us ? = scholars, I guess)]*
Quality control is important.

Humanistic engagement has become increasingly active. In N-america: humanities are under attack as allegedly not contributing to society. including economically (!)
-- how is this different outside NA (apparently not)
- reintegration with public social sphere is integral towards creating a community-relevant aspect of humanities

- Digital Humanities have evolved into a "meta-discipline" but hinders stake-holders.

we should spend more time to modelling to share these models, less writing more making. Contribute to society through techno-culture.

DISCUSSION :

- Is this a sensible approach?
- Knowledge structures: coding things, onthologies, databases. Is it a threat (partly), can it yield some value? Afk vs online: new ways of owning, control, dissemination.
- What is missing? (one of the problems seems the lack of authority? -> shift to facilitating transfer of knowledge)
- How might we best emend (inevitable) elisions?

What We are Doing Now (Tools, Projects)
Content: User-derived content is another imperative for social knowledge creation in electronic environments.

i.e. - the changing of Wikipedia pages and how this reflects a personal engagement with subject matter: who is the expert?

EXAMPLES

- A data portal platform, CKAN fosters user-generated content with its guided publishing process, metadata customization, versioning, and analytics, while allowing users to engage with this data through searching and tagging data sets, social networking, building extensions, and API access. http://ckan.org
- *Omeka* is a comprehensive open source content management system (CMS) for displaying online digital collections of scholarly editions and cultural heritage artifacts. http://omeka.org
- *PyBossa* is a micro-tasking platform that utilizes crowdsourcing in order to carry out small, user-derived tasks and contributions.
http://blog.okfn.org/2012/06/08/introducing-pybossa-the-open-source-micro-tasking-platform/
- Using Mozillas *Open Badges*, an alternative credential-granting system, *Badge Stack* is a WordPress plugin that facilitates the creation of rewards- or achievements-based environments, with earned badges presentable across the web. http://openbadges.org
- Wikipedia http://www.wikipedia.org

Annotation: The affordance of collaborative tools disperses traditional annotative authority.

EXAMPLES
- The extendible AnnotateIt/Annotator allows users to make comments on texts and supports tags, mark-down content, and individual permissions per-annotation. http://annotateit.org
- Co-ment is a text environment designed for commenting and annotation that integrates with multiple content management systems through an API. http://www.co-ment.com
Digress.it is a collaborative plug-in designed bring the practice of marginalia making to the WordPress environment. http://digress.it
- An extensible web application, Domeo uses the Annotation Ontology (AO) RDF framework in permitting fully automated, semi-automated, and manual annotation on HTML or XML documents, as well as both personal and community annotation with access authorization and control. http://swan.mindinformatics.org
- Marginalia, which is capable of integrating with learning management systems like Moodle, includes forum discussion functionality while serving as a straight-forward tool for personal and collaborative annotation. http://www.moodlenews.com/2011/a-gem-of-a-tool-for-english-teachers-marginalia-forum-web-annotation-tool-for-moodle-1-9-and-moodle-2-0/
- TEXTUS is a collaborative annotation platform for text collections which supports bibliography creation and citation. http://textusproject.org

Marking, Tagging, Bibliography: Community bibliography applications, which often incorporate folksonomy tagging, facilitate the collaborative creation, organization, citation, and publication of bibliographies.
EXAMPLES
- *Bibliography Module* for the Drupal CMS, also called *Drupal Scholar* (https://drupal.org/project/openscholar), enables users to manage, present, and cite scholarly publications in various citation styles.
- *BibServer* (http://bibserver.org/) is open source software for publishing and sharing large bibliographic collections on the web through a RESTful API and JSON format, providing also a range of data visualization options.
- Supporting multiple administrators and users, *Document Database* is a PHP-based document management system that uses the BibTeX format in its support for bibliographical data.

- WIKINDX (http://wikindx.sourceforge.net/) is described as a "Virtual Research Environment" and enables collaborative text and bibliographic data management, creation, and use.
- Zotero (https://www.zotero.org/) is an open source reference management system that integrates with word processors and other writing environments. Users can assign tags, organize research material into collections and sub collections, and create topical collaborative research groups for sharing and discussion of libraries and notes.
- *Endnote* -- with new features, has more social elements (Siemens mentions lawsuit between Zotero and Endnote)

Analysis: A growing number of online tools create new opportunities for collaborating during the text analysis process.
- *CATMA (http://www.catma.de/)* is a web-based text analysis and literary research application that allows scholars to exchange analytical results online. The application boasts a number of features: users can apply analytical categories and tags; search the text using Query Builder; set predefined statistical and non-statistical analytical functions; visualize text attributes and findings; and share documents, tagsets, and markup.
- *Highbrow (https://osc.hul.harvard.edu/liblab/proj/highbrow-textual-annotation-browser)* visualizes the density of scholarly annotations and references within individual texts, videos, and audio recordings; it is also able to show groups of analyzed items together in a collection to facilitate pattern discovery by the researcher.
- *TAPoR 2.0* (Text Analysis Portal for Research) (http://portal.tapor.ca/portal/portal) presents a large collection of textual analysis tools for scholars and researchers.
- *Voyant (http://voyant-tools.org/)* is an online text analysis environment (with API support) that allows users to submit texts in a variety of formats to show word frequency, usage, and their placement throughout documents.
- *WordHoard (http://wordhoard.northwestern.edu/userman/index.html)* is a Java application for tagging and annotating large texts or transcribed speech, that offers analysis by word frequency, lemmatization, text comparison, and custom queries.

What tools or type of tools are missing?
Some Current Issues to Consider : ??
- Github . (<-- trying not to think of British slang here) Gimp help? What did he just say?? Gidhelp? --> https://github.com/ (oh) :-)
A DH discussion about the use of github in our field :
http://digitalhumanities.org/answers/topic/how-ready-are-dhers-to-use-github-for-non-code-projects
(It's like using google docs or framapad, but with a much much better "history" structure)
https://twitter.com/frederickaplan/status/350193307588108288

Who constrains knowledge and how?
- By employing digital humanist methods, are we excluding those who are less technologically connected?
- Gamification
- Scholarly engageement : consider the benefits/pitfalls of gamification : ??
- Open journal projects : http://pkp.sfu.ca/?q=ojs

Case Study, Narrative
Social Edition of the Devonshire Manuscript (BL Add MS 174692)
http://en.wikibooks.org/wiki/The_Devonshire_Manuscript

Retooling of our projects after classic/academic way of doing things because people got interested in

"@**RayS6** describes the Devonshire manuscript and its circulation. #**EarlyModern** #**SocialMedia** reminds me of http://www.nytimes.com/2013/06/23/opinion/sunday/social-networking-in-the-1600s.html?pagewanted=all&_r=2& …"
https://twitter.com/radusuciu/status/350196192103235584

Principles of this kind of social edition :
- Process oriented rather than focused on the finished object
- Shares values with Web 2.0: user-driven, perpetual beta, networked.
- Publishing isn't the end of the writing process
- Systemising Shared Authority (chart...)

Our plan was initially more classical academic work, but then...
- To build a peer review in our social editing system, became much more social than thought.
- people were happy to conribute
- Some contents was improved ! (not only typos, but also missing/unattainable content such as data on metioned persons and places. but also counter-vandalizing)
- People were credited and we knew where they were coming from. They were event more engaged then the research scientist in the project !

Conclusion
Models are content for data
key word: interactions, sharing, contributing, collective work
Knowledge production has always been messy!!
not a lot of tools convergency
NB=>systematising shared authority: big deal!

[Course] David Berry(@berrydm), DH and cultural criticism

Meme: http://cartonomics.org/wp-content/uploads/2013/01/i-do-digital-humanities.jpg
Thinking a "critical" digital humanities
Developing a "humanistic" approch to computational phenomena
DH appears to some observers a "Managment friendly discipline"
The digital collections/publications etc. has been accepted as a means of research but not as research itself.
Technician are no longer considered doing the "dirty work"

- dichotomy between academic and technical humanities...

Catherine Hayles : renaming/rebadging/rebrandig to the term Digital Humanities was meant to signal that the filed emerged from athe low prestuge statut of support serivvce into a genuily intellectual endeavour with its own prof practices, rigourous standards, etc. (http://press.uchicago.edu/ucp/books/book/chicago/H/bo5437533.html ?

Re-branding trend: Different research centers renamed as Digital Humanities. "Because it is good money" - says Berry.
(Example in Switzerland: Imaging and Media Lab in Uni. Basel --> Renamed in 2013 "Digital Humanities Lab": http://www.dhlab.unibas.ch/index.php/en/

Steven Ramsay talks about
- DH Type I : TEI consortium + Association for Litt & linguistic computing + Association for Computers in the humanites + Consortium for Computing in the humanites in the early 90ies
Community level = Humanities Computing
including....(examples of practices : encoding of corpuses...ETC.

- Type II : new comers
incl. media studies, people who did things without calling them so, some digital artists, digital pedagogy
http://stephenramsay.us/2013/05/03/dh-one-and-two/

David Golumbia : narrow definition of DH, "DH" with capital letters, "tools-and-archives" as some say (see http://www.uncomputing.org/?p=203)
The big tent definition includes the plain litteral meaning of the term "dh"
Follow them on Twitter, he sometimes argues with Ramsay :)

Why does this matter ? What is at stake ?

Axe I Calculation
- We are not talking of individual act of calculation. But more of this comportment (?) : large scale, calculative, enumeration, representation
- Organisation of knowledge epistemologically grounded in Mathematics
- University becomes de facto buisiness establishments, and much less a place of "contemplative reflexion"
- Digital is hegemonic
- Twitter as a backchannel where people are debating in RT what's happening.
- Heavy focus on tech in DH, discourse & also in projects
- Calculation as a mentality, a "progressive" discourse, as opposed to the "old fashioned" traditonal humanities. Not necesssarly articulate, but implicety accepted./
Implicit distinction in DH with "old fashioned" DH - Both real and imagined dichotomies
- Traditionally, outputs of research like code, archives, tech reports are not taken into account when it comes to publishing.
- These dichotomy are both real & imagined ; productive & improductive
- Often the outputs of DH are not yet conceived scholarly for tenure => generates precarity
- Keen early adopter mentality...
- Journalists love DH, in contrast to their lack of interest in traditional Humanities
- The term DH is new-ish (2001)...
- Debates over XML/TEI = public expression of "Digital formalism" that many DH projects often contain
- The problems/notions of acronyms: "you're not really doing a digital project if it doesn't have an acronym" --> technical language changes the way we frame/formulate the project
- Ref: CULTURE MACHINE VOL 12 • 2011 www.culturemachine.net • 1 - *THE COMPUTATIONAL TURN: THINKING ABOUT THE DIGITAL HUMANITIES* (David M. Berry)

- Bulding things, "more hack, less yack" = very strong in DH community (cf. Ray Siemens rhetoric) = construction & making of digital systems, archives, interfaces, visualisations = PROGRAMMING = code-focuses
- this has been seen as a lack of critical engagement
- "Making things" as a marketable skill
- Detractors of DH see DH as a move by neoliberalism/mangement within the university
- Project-oriented mentalities

Axe II Organisation
The university is a methodological commons of computational techniques
- Is this a kind of "scientisim"? The humanities taking science as its model
-" Putting aside the question of the ‘‘two cultures,’’ we nevertheless observe a striking, non-trivial resemblance between humanities computing and experimental science: both are data-centerd, equipment-orientated activities
that centrally involve modeling and tend to be collaborative.", McCarty
http://www.mccarty.org.uk/essays/McCarty,%20Humanities%20computing.pdf
- Risk of seeing the computer as "truth machines"
- Constant struggle for funding
- Shift towards research
- Reading the DH as a revelator of how the Academia is evolving in a neo-liberal economy
- Universty: risks to become corporations, professors project managers.
- Input/Output calculation: enure that capital invested is repaid in termes of knowledge produced and disseminated
- DH is a savious AND a sinner in relaiton to the future of university
Grusin: http://www.c21uwm.com/2013/01/09/the-dark-side-of-the-digital-humanities-part-1/
Stanford literary labs -- Pamphlets: http://litlab.stanford.edu/?page_id=255

Axe III Speed
mechanical increase of speed with the power of computers

Transience becomes an important aspect
These new technologies enable team-based working methods, offering the possibilities of new social ontologies in the Humanities: ‘Big Humanities’ – not necessarily collegial and flat.

Just in case: possible amendments can be made afterwards as well, David Berry gave a previous talk on the issue http://vimeo.com/62788322 and later on infoclio.ch

What are Critical DH ?
- How political views are embeeded the in tools themselves
- See TransformDH Group 2013: http://transformdh.org/

- Productive slowdown forcing a project to reflect on its approach, method and goals (--> slow HD?)
- critical thinking acts as "grit in the machine" (though DB is looking for a better metaphor)

QUESTIONS/DISCUSSION

EP - Digital Humanities starts with a research question; yet there's a proper recognition of technology as a strategy. Technology is considered a given...
DB clarifies that after the question is phrased, you use technology to structure the project

Algorithms and their importance for the digital humanities. Important questionmark over the algorithm

- To what extent is DH a response of governmental demand that humanities(?) justify what they do?
DB suggests that while DH language fits in nicely with justifying activity in "management speak", the DH is not a cynical exercise in working a system. It emerges out of a more authentic space of engagement with humanities issues.

Abigail de Kosnik's work on 'free labour' in the digital enviroment may be of interest here: http://spreadablemedia.org/essays/kosnik/#.Ucw49vn7AZg

Tutorial/Workshop Session 1: Elena Pierazzo (@epierazzo) : Digital Textual Editing [notes by @railyards]

15 minutes theory to introduce XML and TEI.<== Did it work ? yes ! \o/
Tool used is http://www.oxygenxml.com/xml_editor/tei_editor.html - 30 days free trial available

- TEI, a standard for text encoding.
<TEI xmlns="http://www.tei.c.org/ns/1.0">

- <teiHeader>
- METADATA HERE
- </teiHeader>
- <text>
- TEXT HERE
- </text>

</TEI>

Other types of elements, with different "functions": <sourceDoc>, etc... See http://www.tei-c.org/release/doc/tei-p5-doc/en/html/REF-ELEMENTS.html
Metamark : indications written in the text.
Gap : when there is something we can't transcribe (read). For instance, <gap unit="chars" extent="1" reason="covered with ink"/> means "there is something we can't read that is 1 character long and we can't read it because it is covered with ink

Class is divided in two: Beginners and intermediate. Both group must transcribe a manuscript by Austen, dealing with special chars, etc.
1.
2. http://www.janeausten.ac.uk/manuscripts/pmwats/b1-4.html

Tutorial/Workshop Session 1: Lynne Siemens (@lynnelynne53?): Collaborative Work Practices in the Digital Humanities [notes by @squintar, @radusuciu and 2 more]

Questions from the participants :
How to form a team?
Project Management ?
Internal/external partners ?
Technology to support team
How to engage people ?
Communication across discipline ?
Integrate new members ?
How to scale ?
Team skills
How to establish collaboration and culture of collaboration?

Questions to be asked before beginning
- What are the project's research question & objectives ?
- Are additional people & skils needed to accomplish the goals ?
- What kind of relationshiops is desired between collaborators ? Contract ?
- Team is a set of individuals who interdependently & jointly accountable for outcomes. Has to be at least 2 people ;)
Project's research question decides the team and relationship between project members.

Benefits : Work & responsability is distributed. Helps having higher quality ideas thanks to diversity. Social & fun.
Challenges : Time consuming (scheduling, understanding the question, etc.), Coflict management & coordination, New skills needed
Conclusion : not for everyone/every project !

Does one really need a team?

What skills are needed ? (technical, content, collaborative)
What individuals are needed ? Are they available ? Are they collaboration-ready ?
Early definition of responsibilities and roles => WHO will do WHAT, WHEN, and to what LEVEL of quality ?
Team members can be needed for reasons other than research e.g. funding purposes etc.
There is no right team size: depends on the project

How to pay: depends upon what best is required by a team member? There is pay and pay: money or credit or academic recognition.
In fact is project management recognized in academia as something useful. Not really... only when your project shows considerable results and get more funds Ok but even then sometimes that fails to be recognized by academic or tenure committees. Cf. this blog post by Sean Takats of CHNM: http://quintessenceofham.org/2013/02/07/a-digital-humanities-tenure-case-part-2-letters-and-committees/

"To recap: Conceive projects? Service. Develop prototype software? Service. Write successful grant proposals? Service. Write code? Service. Lead developers and designers? Service. Disseminate the results of the project? Service."

How is a research team formed ?
Stages : **Forming** => **Storming** (??) => **Norming** (guidelines, how is the team will work together, time to respond to emails, frequency of meetings/skype; authorships) => **Performing** (things getting done)(sometimes nothing gets done, loss of resources)
Norming should clearly indicate about academic credits and authorship
Once a team works, you want to work with them again, without cycling in forming/storming/norming phases. Except if a new person comes.
Suppport the team and sustain it. Supporting the team will help sustaining it
Ensure performance, member satisfaction, stakeholder satisfaction
Team building & sustaining : over time, distance, institutions, disciplines. Trust & communication.

How does one ensure research effectiveness ?
Plan in advance inc. good meals
Use multiple forms of communication
Create small successes upon which to build
Remember this takes time
Document agreements
What do we do in case of miésavente

What are some tools to support resarch teams ?
Match the pepple : Responsability, available, when, etc.

Le logiciel qu'elle conseille est Basecamp, mais aussi Google doc, iCalander, google spreadsheets, etc.

[in the margin : wondering when it's gonna be about DH] :) I hope so ,-) TROLL ! :o) c'est quoi troll, i feel ooooold, Je te dirai si tu dis qui tu es ;) look at the list, match the colourvoilà I wasn't talking to you but to the turquoise guy :) I know, and I know him. And he is a big troll. And I'm two seats on your right. i've been PRISMed

Tutorial/Workshop Session 1: David Berry(@berrydm): Digital Humanities: Critical Approaches in Digital Humanities [notes by @xaentenza]

EXERCISE: Start thinking about Digital Humanities. David Berry starts asking: what comes to mind? The key to notion of DH? (keywords given by auditorium)
- sharing
- community
- interactivity
- computation (digital)
- archives (clearly a keyword, and what he calls "infinite archive")
- visualisation
- funds
- cybernetics
- search infrastructure / environnement

- Virtual Research Environnement
- national culture
- iteration
- Open Access/copyright
- discovery
- digital literacy
- databases
- remediation
- access: TOS, Licencing, Use
- curation
- big data
- editing (text editing)
- publishing
- tools and methods
- interoperability/DH Islands
- Crowd sourcing
- "More hack less yack"
- gamification
- building
- transcription
- academic promotion, recognition
- GIS, mapping

Map of keywords associated with DH on blackboard:
https://twitter.com/xaentenza/status/350277297682644993

EXERCISE Choose 5 keyterms and think how they play out (groups of 6 persons).
Every group gives now a summary of what they talked about

[Bad wi-fi, and internal discussions between the groups hard to reproduce]

Group 1:
digital photography changed last years. Photography now is about stream. Got photographs by systems.
Argument about "remediation" Remediation, an overused word? A crouch people use, empty word. But for others, remediation is useful (but why).

Group 2:
Talking about gamification. Gamification: difficult subject. Gamification is using of game structure, encouraging using tricks to get the users to do some things.
Good things and what is just "sexy" things for DH. Talking about cheating the users.

Group 3:
Talk about transcription and its problems. In Lausanne, program building database about Pays de Vaud in XVIIIe century (Lumières Lausanne, http://lumieres.unil.ch/). First, transcribe documents in Word by instruments. They noticed is that some students wanted to reproduce the words layout. Also discovered that some didn't know about writing in the 18th Century. Lot of work to edit and correct all of that word data.

Group 4:
Began with sustainability and talked about digital data, archeological missions.

Group 5:
Practices in different countries, promotion, libraries, ...

Tutorial/Workshop Session 1: Nicolas Chachereau(@nchachereau): Zotero and Citation Management Softwares [notes by @twinkling_sheep here]

- introductions of everyone here (very nice group of diverse people!)
- demonstration of the manual pitfalls of referencing by hand as opposed to the ease of Zotero in footnotes
- discussion of some of the other reference options

A List of Possible Citation Sources

- - Endnote is well-known (license is offered in many universities (i.e. standard in Lausanne) but as soon as you leave university you have to pay for upgrades): http://endnote.com
- - Mendeley (free): http://www.mendeley.com
- - Litlink: http://www.litlink.ch/home/
- - Citavi (free if you have less than 100 references - not very useful for a dissertation): http://citavi.com

Why Zotero?
- easier than endnote?
- free and centralized
- reference to Mendeley's comparis: http://www.mendeley.com/compare-mendeley/
- problematic with Ipads
- way of collecting, organizing, citing, synchronizing, and collaborating
- also free for downloading
- accessible both online and on laptop (can sync) although pdfs won't be available

- -- additional storage space can be bought, but you can also hyperlink to jstor articles etc.

- two versions of Zotero: one is a stand-alone; the other is a firefox extension. Recommended that you use Firefox extension version, esp as you can link it to the way you search and because it was originally designed for Firefox (nevertheless i have downloaded it to Safari). But it is equally possible to do it through Safari.
- personal downloading tutorial ensues :)

Once downloading is complete, we go to tiny.cc/CHDH-Zotero and the "real" tutorial begins!
- open Zotero application, which should install the extension directly onto your browser (so you click the icon to save the reference) --> must also install the browser extension

Some tips:
- sometimes when you click, the entry might appear in lowercase (or similar): you can right-click, transform text, and then change the title case or sentence case so that things will get capitalized automatically
- manual additions possible: click green "+" and add entries automatically
- Zotero isn't the ideal method for citing archives/manuscript
- when you save off browser, it'll automatically download the pdf as well
- make changes to entries as you go along (make sure your import looks like and represents what you've imported)
- *Troubleshooting*: the option to save the snapshot from Zotero is not available automatically on Safari (only on Firefox). To do so, right-click anywhere on the screen and press "Save Zotero Snapshot from Current Page"

Adding a citation to Word (as footnote):
1) - click Zotero insert citation button (r."z) (not ideal for MLA style [discovered from experimentation])
2) - choose style
Note: Zotero.org/styles – can download styles from Zotero from here
3) - click on book of choice that shows up in library of references
4) click on bubble that appears after selecting book and add page information
- can add footnote as well in MLA-style
--> select footnote from insert menu, then put in additional information to "prefix" or suffix option (no word limit)

- possible to export Endnote library with formatting etc.
* wiki on Zotero page that gives all information about this (in English)
- export to RIS and then import to zotero (double check when internet goes back up)

Suggestions from Experts
- go to Zotero website and create an online account: then create login and password to synch data
- Back up Zotero library all the time!!
- go to "Preferences" in Zotero, advanced tab, and create a dissertation file (or whatever the title of your research project) so that it doesn't get lost in the random mozilla or safari file.
- use "tags" to organize bibliography (tags will appear in hyperlinked words at bottom of Zotero)
Note: import from library catalogue: will create tags as a default (you can get rid of it by going to general preferences and dis/enable that feature.

To create a bibliography without making it a citation
1) select references
2) right click on selections and select "Create bibliography from item"
3) choose style (MLA, ALA, Chicago, etc)
4) choose output mode (i.e. copy to clipboard if you're making it into bibliography on word document, and copy/paste into word doc)

- can cite from Zotero in an email:
option 1 – drag and drop
option 2 – same as above for creating bibliography

_____
thank YOU for contributing! (keep it up)
_____

-----Friday 28 June 2013

[Course] Claire Lemercier, Quantitative methodology & Network analysis

Personal page: http://www.cso.edu/cv_equipe.asp?langue=en&per_id=168
[download her presentation slides here: http://tinyurl.com/oxk8kdr]

hello everybody \o/ Salut tout le monde !
she has an old phone = a proof she's serious ;) [this text is totally unrelated and should be deleted]

Network analysis seen as a toolbox : she's gonna discuss the "fashion of network analysis"
In fact, what is a network perspective ? what can it help you to see ?
Less is more : pour rendre visible qqch, + c'est complexe, - c'est utile.
En analyse de réseau, les outils ne sont pas complètement distincts des questions de recherche = Most important are research questions, not the tools \o/ Yes, what she said is that the choice of the tool will depend on the perspective/research question. A tool is not neutral.

[sources : website http://www.quanti.ihmc.ens.fr/
her book : http://www.amazon.fr/M%C3%A9thodes-quantitatives-pour-lhistorien-Claire/dp/2707153400/ref=sr_1_1?ie=UTF8&qid=1372403372&sr=8-1&keywords=claire+lemercier
the book has been translated into English, but she's still looking for a publisher
she also cites the "SAGE handbook" : personally I think it's not a good book, especially when beginning with network science]

Terminology: "Formal Methods" rather than "Quantitative methods"
Not necessarily big data, but deep/dense data, even if the dataset is small. The method is about dealing with complex, multidimensional data set.

Counting, but mostly correlating and finding patterns. (Correlation with music) Not necessarly "macro", but changing scales. <--> Data Mining
Linked - and not opposed to - with qualitative methods : no dichotomy, frontiers blurred

Changing scaling is important, from middle level to individual levels, and macroscopic level (zoom in/zoom out)
Zoo of methods, we need a map to navigate into this space
Source: Tilly, 2004 <= did you find the complete reference ?
She shows a typology of network in social science and which approach you ca use depending on the use of network as a descriptive or explicative variable.

Distant and close reading: Lemercier and Mata, 2011

http://hope.econ.duke.edu/sites/default/files/Mata%20and%20Lemercier%20-%20Economics%20at%20Newsweek-formatted.pdf ??
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1753164

Other ways to look for patterns and navigate scales
1) sequence analysis: technique to explore trajectories (in time, something happening to a long text, recurring patterns, etc). Eg context of narrative sociologie.

TraMineR package for sequence analysis
http://mephisto.unige.ch/traminer/download.shtml
http://cran.r-project.org/web/packages/TraMineR/

- can zoom in and out on trajectory patterns as a whole, but also can explore exceptions

- need to find the methods that suit your needs: network analysis -- the peculiarities of it suits humanities

http://www.sciencedirect.com/science/journal/02613794/31/2

Network Analysis

- Social Network Analysis (SNA)
- Has become a research tradition
- Shared vocabulary (despite potential of extremely different theoretical paradigms)
- Network analysis is not a software, not a theory
not married to life to any theory/shool of thought (Bourdieu, Latour), just a way to look at things
- A tool is never neutral!
- The novelty with network analysis is the focus on relational data ("ties", "relationships") instead of only using attributes of actors.
Focus on relationships
What has been developped for a use in social science can be ??

- "Anything can be considered as a network" : not specially how people tweet/like/etc. each other = Ties and non ties between anything (people, images, texts)

- Some networks do not require too much analysis, some are just lists (if all relied to the same point, and no connection between them) = No added value of presenting a list of contacts... Doesn't add analytical value
- Network not a group but a set of ties.
- Ties are interesting because there are not ties everywhere (if there are everywhere, it become less interesting)

- General idea is to find structures/patterns and to this at different scales: the great strength of this method
- Don't stick to the macro view, do zoom in = navigate scales !

- NetDraw (http://www.analytictech.com/downloadnd.htm) = this software is the anti-gif..!
She advises us to begin with this software, even if you use Gephi after [funny : I would advise never to use it]

[one thing for sure : netdraw outputs are ugly] (but we can decide not to bother on the aesthetics of the stuff) [if you don't plan to publish or present it ok]

- Sometimes (picture anybody ?), there is no structure but a lot of connectivity, because of lots of hierarchial (?) links
- "Structure is created by absent ties" (read Burt's structural holes articles on this subject, for example http://books.google.ch/books?id=E6v0cVy8hVIC&redir_esc=y)

Motivations for using this:
- a fresh look to look at your data
- increased ability to recognize patterns (strucitre in a mess or borders... created by absent ties)
- Get a feeling of what happens to my data, first approach, when you are alone with your data.
- Teach/show things. You can also hide things and look what happens.

Floor is opened for questions on this first part:
1) Is it possible to have visualization in 3 dimensions with NetDraw? --> ha ha nice. I've been trying to figure out the name she was saying (being a total digital pleb)
A - requires other software, not with NetDraw

Interesting: to think at different ways to look at the interactions between data
- you have to have a matrix and figure out in each case what you're trying to achieve: the idea of being systematic helps you to get out of a tautological rut

Thinking about ties
-Not "is there a network there ?" : there is always a network !
- Look at the ties : two hypothesis :

- - what ties do ?
- - where the ties come from ?
- - How to think about defining ties?
- - don't put all the ties in a big bag, you'll get only trivial tings

Networks from Humanistic Sources : some examples
- it's not only for sociologists, historians can also use them
- lots out there (no dearth of sources)

Examples from history
1. Records of interaction : who is recommanding who (cooptation)
2. Account books => Pierre Gervais, 2012 from Gradis accounts (transatlantic trade)
3. Citation and co-citation : for history of science/knowledge => already done in the 19th century (chemistry) , citation in letters, citation of places

Example from Litterature
Mental maps of contemporary poets (Dubois & François, 2013) : poets mapping their colleagues in anthologies and by labels

Example from political sciences (?)
Tilly, 1997, Structure of petitions to the Queen (1758-1801 & 1832-1834), with the words used "attack",

Example from anthropology (?)
Folk songs of/about miners, studied by Marion Henry (Iramuteq was used http://www.iramuteq.org/)

Stolen from David Berry's tweet: Digital humanities? for old profs don't understand tech, & young ones don't understand hermeneutics? **#dhch** http://lareviewofbooks.org/article.php?type=&id=1801&fulltext=1&media=#article-text-cutpoint …

Ref/example: Sigrist and Widmer 2011 - Network of relations between botanists in 17th-19th century - Nodes are placed on the graph according to their date of birth. Full text: http://revista-redes.rediris.es/pdf-vol21/vol21_7e.pdf

Tutorial/Workshop Session 2: Claire Lemercier: From Sources to Databases: Data extraction out of Humanistic Sources

Complicated relations are nothing more than rows and columns
- please do not become slaves of programs; the basic structure of data is in quantitative columns
- will not be discussing software, but beginning with some sort of material that is accessible
- underdiscussed part of research: inputting?
- put data in rows and columns
- transform data as literally(?) as possible
- "code data" not in a sense of programming but in terms of signifying
- separate "inputing data" and "coding data" (become prisoners of own coding)
- flexible ways of inputting data that keeps open the possibilites of dealing with data but still adding structure
- call for people who have data and how we would input or code it according to the (negative) principles she'd like to show us (principles of what we shouldn't do)
-- this notetaker is confused about what she means by material... :(

Case Study 1: Caravanserais

Criteria:

- - name
- - geographic coord.
- - cited/discovered by (first name, surname, source title)
- - material (stone, brick...)

- - date of construction
- - water installation
- - roads/paths
- - archeological data (+/- details)
- - who built it
- - near to village?

(ca. 72 caravanserais, Syria)

General aims of building the dataset:
1) She is trying to see the development of the building techniques, so her interest is in the material and archaeological results, reconstructing patterns of patronage (based on who commissioned the road)
2) reconstruct the road networks on the basis of caravanserais and stopovers cited by different travellers
3) see the caravanserais distribution patterns
Problem in humanities: incertitude!

Question: if you have several sources, how do you deal with it in data extraction?
L'entité sur laquelle je travaille est-elle si évidente que ça? It is an option to focus on the source and not on the unit (here a caravanserail)

It's not because it is a table that you have to be 100% sure about your data -- shouldn't you always be 100% sure about your data?? you can also express your doubt (for example, datation: year 1455, 15th cent., uncertain, unknown, ...)
[clarification: you should be sure about your data (what you've recorded) but not about its exactitude and relevance/usefulness to the research project as yet (?)]
- should use software that is flexible but allows for retrospective notes/columns/changes, etc.
- if your data is more structured, you should always be able to transform the data into another table

Comment séparer les informations qui sont miennes ou qui viennent d'autres collègues/études? Il est possible de faire des "doubles colonnes" par ex :

DATE	
1455	because I found it in this reference...
1472-73	depends on the day of birth of X

Case Study 2 - Italian politician (?):
Heterougenous documents written by one person :
- type
- date
- opinions -> concepts (closed or open list) How? Most frequent lemmas

General aims of building the dataset:
- Have a better understanding of the chronology of ideas by the studied politician

Case Study 3 - Arabic versions of the New Testament:
How to deal with uncertainty ?
- It's better not to have empty cells.... Specify if unknown, or if inexistant, and why it is unknown or inexistant. (always good to explain why you have no data - not just for statistical purposes)
- Il vaut mieux avoir des ; que des espaces (reconnus par excel)

Recommended software:
- pajek
- "clustering" is a family of techniques used to create typologies according to similarities

- - R is the software that helps map this (see below)

- - beginner's "how-to" tutorial offered by Lemercier in another forum (she will send us PDF later) That would be good too for those of us who couldn't attend this workshop.

- family of algorithms called 'clustering" not with excel or calc but with statistical software called R (http://www.r-project.org/)
- TraMineR (see notes from Lemercier's lecture)
- package that provides you with something to click: RCommander (the most used statistical collations)
- additional package for correspondance analysis "FactormineR" -- allows you to export your data back to where you're more comfortable (http://factominer.free.fr)
- alternatives to R - SAS, Stata, SPSS [don't follow that recommandation, please] <-- it wasn't mine! I'm just recording!! it's all acronyms to me..! ;)
[that's why I'm putting it in-between brackets : it's my opinion, I'm not even present at this workshop]
(hi, it's Claire - I said you should definitely use R (several times) but if it happens that in your University you have commercial software and nice colleagues who want to work with you on it, well, why not)

Tutorial/Workshop Session 2: Susan Schreibman (@schreib100): Text Analysis with online Tools @squintar

Tools we will test
1. http://dfr.jstor.org
2. google Ngram viewer : http://books.google.com/ngrams (You cannot add your data to it, it will work with the scanned books only, mainly for comparing how particular words' usage has chnged over the period of time in the scanned books.)
3. Text Arc : http://www.textarc.org/ Brad Paley Developed "for fun" by designer, independant from academia. Based on concordance (?was she talking about this soft?) Get an overview of an entire corpus of texts
4. Wordle http://www.wordle.net/
5. IBM's ManyEyes : http://www-958.ibm.com/software/analytics/manyeyes/ micro-reading of texts, charts out the words that are led by certain words
6. Voyant : http://disc.library.emory.edu/lincoln/voyant/ developed by literary scholars. shows you also where in the text the word occurs. http://voyant-tools.org
7. Data for research : http://about.jstor.org/service/data-for-research

Other tools and Resources

8. Rapid miner http://rapid-i.com/content/view/181/
Bookworm Culturomics
HathiTrust

First step : chose a text ;) We'll each play with one single text (personal, not too short) with several different tools
-- how's that going?
.doc prepared by copy pasting the text of a PDF of a major paper in biology/medical sciences (first paper of discovery of HIV before it was called HIV) (as in Grid?)
Trying to do it with Unicode text, Wordle is not able to render it properly, just some squares and all....

My work in progress can be followed here :
https://docs.google.com/document/d/1oQVt1w0OEHdBAMkvhW0lj-7opDGCnFi5xVB88-lGRaA/edit?usp=sharing
Does these things work with languages written using scripts other than Roman ?

Using ManyEyes with french .txt is kind of frustating because it doesn't understand utf-8...
With Voyant it is much better. With Voyant you can add your own limits, stop words list, which is usefull.

Tutorial/Workshop Session 2: Claire Clivaz (@cclivaz), & Lukas Rosenthaler: Multimedia Literacies and SALSAH

Towards Multimedia literacies. Discovering SALSAH (System for annotation and linkage in arts and humanities). http://www.salsah.org
Part I - Claire Clivaz
Swiss institutional context about Reasearch Data Management in the Humanities:
http://www.sagw.ch/fr/sagw/laufende-projekte/ddz.html
IPads and Digital Tool in elementary classroom -- > Beginning of a new era of digital literacy, says Clivaz
1860-1980: Generalisation of elementary education and monopole of the print culture
In Antiquity, "sound-image-text" culture. We are too much formated by text-centric culture.
Emergence of Multi-media Academic Virtual Objects
Project example: www.zeega.com - Artists meet scholars

FYI: The clip of Youssoupha, Menaces de mort, was about/against Eric Zemmour (not Nicolas Sarkozy). http://rapgenius.com/Youssoupha-menace-de-mort-lyrics#lyric

Part II - Lukas Rosenthaler
Traces the history of the Digital Humanities Lab in Basel:
http://www.dhlab.unibas.ch/index.php/en/
First digital camera in the Lab in 1981
Today many online primary sources available.
Project example: http://www.e-codices.unifr.ch/
If you want to work with digitized sources, you have to print them out or save them locally, you cannot annotate them, i.e. you miss a Virtual Research Environment (VRE).

The web was founded as a virtual collaboration tool for CERN researchers. And then it has developped in... well, you know.
Why digitize large collection if you cannot work with them.
VRE for Humanities. Features: Annotation / Linkage / Transcription / Collaboration

Technical aspects:
- Resource Description Framework: model developed by the W3C (RDF, semantic model)
- Salsah supports bi-directionnal connections between databases
- "We are very flexible and we can added practicaly on the fly".
- Web-based frontend appliation <- html5 / ajax / jQuery
- Flexibility: possibility to import metadata based on DublinCore and manage them
- Bi-directionnality RDF to RDFS (cf. schema http://en.wikipedia.org/wiki/File:Semantic-web-stack.png)

Search strategy: different ways for searching (text, graph and... ? Didn't catch the rest)

Image store:
- high quality (JPEG2000, lossless compression), conversion (JPEG, TIFF, PNG) => possibility to adapt the quality of images in reading
- optional watermark

DEMO

Tutorial/Workshop Session 2: Laurent Pugin & Claudio Bacciagaluppi: TEI and Musicology @yrochat + @xaentenza

[Note: Laurent Pugin & Claudio Bacciagaluppi are not from Bern University :-) Members of the Swiss RISM Office http://www.rism-ch.org/]

1) How is TEI applied to musicology and reseach?
2) MEI (correspondace of TEI, Music encoding instead of text encoding, I for initiative)
3) Short presentation of our interests

Introduction to TEI http://www.tei-c.org/index.xml
- Machine readable
- goal is encoding data along with metadata
- Possibilites of defining tags and linking to standardized ref. tags.
- no computer science skills required (but XML)

OxGarage conversion from text to TEI
then editing in Oxygen (30 day trial for free)

Presentation of projects
- E-Codices, from the University of Fribourg http://www.e-codices.unifr.ch/fr

- (In e-codices, the TEI Header is often bigger than the body because there is no transcription, but only metadata about the document)

- RISM : Historical music inventories 1500-1800

- In Mozart's Words: http://letters.mozartways.com/ (correspondance of 1400 pieces)
- Carl-Maria-von-Weber-Gesamtausgabe (WeGA): http://www.weber-gesamtausgabe.de/de/Index (rendition in typography, XML,)
- Kapellmeisterbuch: A musical diary about what was played when http://d-lib.rism-ch.org/kapellmeisterbuch/
- Hofmeister XIX http://www.hofmeister.rhul.ac.uk/2008/index.html

Do you choose to make a database or make transcription in TEI?
Things change with semantic web. Still important to have a transcription of the source, see what was written, take different parameters into account.

Music encoding

Graph: https://twitter.com/yrochat/status/350547779287924736

COMPOSE ----- WRITE

Music / Graphic notation / Neume notation / tablatrure / Scordatura / Score

PLAY------ READ

WAV/AIFF MP3/ACC MIDI/Smoke **kern Music21 MusicXML/LilyPond

Samples Selected sample Events Music Symbols Music Symbols + Layout indication

MEI is the small sister project of TEI

XML-based music encoding

MusicXML (Michael Good, CodaMusic) >> Bought by Finale
- Interchange format for software applications (Finale, Sibelius)
- Deeply inspired by MuseData

IEEE 1599 (Universitá degli Studi di Milano)
- Designed to describe heterogrenous music contents comprehensively (linking between notation and audio)

CMME (Ted Dumitresco, Utrech University)
- encoding format for digital editions of Renaissance music
- Developped in conjunction with an edition software applciation

MEI (Perry Roland, University of Virginia)
- Deeply inspired by TEI
- Community-driven, modular and with focus on philosophy
- (hosted by Mainz academy)

Others: MusicML, MusiXML, MusiqueXML, Xmusic, etc. ("the biggest difficulty for these guys was to find names containing "music" and "xml"")

The Music Encoging Initiative: http://www.music-encoding.org

Why did become MEI it's own community: music notation works in a different way as alphabet: symbols, rules (two notes grouped together with a *b*), etc.

Example of note encoding:
<note pname:"b" oct="4" dur="4" dots="1"/> (un "si" noire pointée, troisième ligne, https://twitter.com/yrochat/status/350550335661670400)
Other example: https://twitter.com/xaentenza/status/350550597784715264
Multiple hierarchies https://twitter.com/yrochat/status/350550704701714432

The line between content and representation is difficult to draw

Multiple hierarchies (a general problem in XML) > Very common in music

Example and sound sample of Baude Codier's music (Codex Chantilly)
http://en.wikipedia.org/wiki/Baude_Cordier

Ars nova- Partout qu'on passe suis composé
The text of the song tells you how to read the notation

"Spiral galaxy" (George Crumb, 1972) spiral musical notation
https://twitter.com/xaentenza/status/350552396742660096

Josquin des Prés (c. 1450/1455 – 27 August 1521) : black notation to significate death ("la déploration sur la mort d'Ockeghem" score and sound sample)
https://twitter.com/yrochat/status/350553046029316097
http://en.wikipedia.org/wiki/Josquin_des_Prez (starred)

"If you want to encode these into something machine-readable, how do you do that ?"
Sometimes you have a notice made to help play the music. How to include that ?

Writing and representing MEI

Aruspix http://www.aruspix.net/

The examples given (some pics on Twitter) were here to remember what is music notation. Now the workshop is moving to applied examples were the scores have been encoded into MEI.

The demo is impressive. You can compare different versions of the same score, and interact with the measures and notes and find which parts appear or don't in the different versions. The score is interactive. It's svg format.

Is there web version? Some project exist (McGill ?)

Typical workflow https://twitter.com/yrochat/status/350556760353955840

Microfilm
OMR in Aruspix
MEI for orignal source encocing editing in Aruspix
MEI for critcal edition eencoding

Rendering music encoding on the Internet

There is no standard way for rending music on the net.

Plug-in for binary files

Plug-in for MusicXML files

Early music specific tools

Javascript libraries

VexFlow
Notefligh HTML5 score viewer
ABCJS parser, editor and reder for ABC (folksonsgs notation)

What makes music rendering conplex (a lot of layers to take into account !!! For example the beams http://en.wikipedia.org/wiki/Beam_%28music%29 depend on the alignment of the notes which depend also on their duration, for example dots, sharps, etc.)

TEI SIG on music http://www.tei-c.org/Activities/SIG/Music/index.xml

MEI and the TEI Genetic Module

Search engine ?
Not one way of searching
Rythme Pitch

Search the LIber Usualis (1961, 2000 pages)
http://ddmal.music.mcgill.ca/liber/
Example : agnus dei
http://ddmal.music.mcgill.ca/liber#type=pnames&query=edcdeee&result=1&p=157

Music NGram viewers http://www.peachnote.com/ (it doesn't work with the pink panther !)

[Course] Frédéric Kaplan (@frederickaplan), Time Machines [How to Build a Time Machine]

[Kaplan's personal web page: http://www.fkaplan.com/en/pag1-bio.html]
Project Website: http://dhlab.epfl.ch

- changes to Geneva (prior it was almost a walled medieval city)
- panoramic view of Geneva - reconstruction of Geneva in 1850: http://www.panoramio.com/?no_redirect
- relief Magnin, at Maison Davel: shows the city in ca. 1850 http://www.ville-ge.ch/mah/index.php?content=6.2.1.3.5.&langue=eng
- using "bird's eye view" as a tool without realizing how revolutionary it is
- What is missing of Google's Maps & Street views is the time factor

- http://robotflaneur.com/ based on google street view thanks to the work of google that made the space machine-readable (ie not only visible) : we now live in a huge map \o/ => did he say so or not ?
- Check this game based on Google street view : http://geoguessr.com/
- we are living in a "big now" - relationship between time and space (high intensity "now")
-
- Hilarious depiction of year 2000 prediction: "en l'an 2000" images: http://www.ufunk.net/en/insolite/en-lan-2000-le-futur-imagine-en-1910-avec-24-illustrations-retro/
- Travelling through space and through time: can we build the tools to do it?

Can we build Google Map of Past, Facebook or other social networking for past?
Information mushroom: http://fkaplan.files.wordpress.com/2013/03/champignon-en-lucida-3.png
Digitization of information available can help with that as well as extrapolation of information (simulation).

Used Digitization and Simulation for VENICE TIME MACHINE:
http://fkaplan.wordpress.com/2013/03/14/lancement-de-la-venice-time-machine/
80km of archives of Venetian history over 1000 years written in different languages, documenting every detail of Venetian history
Goal: Transform 80 km of Archivial documents spreaded on 1000 years in a single digital knowledge system
Digitizing 450 volumes per day! And so 10 years. (First step of the project).
Maturing techniques to improve efficiency
- digitize, transcribe (crowdsourcing), translate, index documents
- demonstrates the development of language and also the meta-development of documents/documentation
- speech recognition software might show a way to improve OCR of historical documents if we know what kind of text we expect
- next step: semantic extraction of 10 billion events for different type of information search from it.
It will allow us to discover how Venice developed as a city as well as an empire
- from this data, we can also extract information about, say, maritime routes (and we can build a Mediterranean simulator)
-- allows for a recreation of maritime habits (i.e. if I were living in 1323, how would I get from point A to point B) ["routerank of the past"]
http://orbis.stanford.edu/

Inconsistencies have to be taken care of while generating these simulations. These include:
A.Incompleteness
B. Errors
C. Falsification
in primary sources;

and
A. Inconsistencies and Contradictions (intra as well as inter document)
B. Errors in transcriptions
C. Interpretive Biases
in secondary sources

Also, some additional ones during processing and pre-processing of documents.

Coding metadata permits to speak about the notion of "fictional space" (in order to say that we speak about possible representations, and we always deal with a "truth").
Metadata coding will not ensure correctness but a procedure for correcting whenever it is found. A kind of inconsistencies tracker.

Coding uses RDF standards and some additional TRICKS....

Problems of the word fiction: he prefers the word fiction because it has the potential of being true

Who is this tool for? propels us into second half of talk
- larger goal is to create a sense of attachment (the VTM is their project)
one strategy: make the search/visualization/contribution tools disappear...
- Aimed audience composed by researchers AND a large audience to engage them in various ways to evaluate using some evaluation matrices.

The missing piece: students, prepare them for a change, which can happen or is seen as happening as it did in life sciences. So that they are ready for the Paradigm shift in the Humanities: large-scale projects

Teaching
Course of Kaplan: dh101.ch #DH101 --> give them the feeling of being part of something larger than the course
The missing piece: students, prepare them for a change, which can happen or is seen as happening as it did in life sciences. So that they are ready for the Paradigm
Guiding Challenges while teaching the students:

MOOC courses as a way of teaching students and engage them in such large scale projects and engage them in a distributed manner where they learn while working on REAL project. (and crowdsourcing)

The project tries to invent a new kind of large-scale academic projects.

Question: What about financial aspect ? Well... it's a scientific adventure...

Question: What kind of team is already in place?
- a big one... but one that always has room for more. ;)

Question: How can you be so optimistic about manuscript OCR?
-

Tutorial/Workshop Session 3: Frédéric Kaplan (Ecole Polytechnique de Lausanne)(@frederickaplan): Semantic Modelling for the Humanities: RDF and beyond

Goal: Feel the beauty and power of semantic encoding

Lot of (ugly) acronyms: RDF, URI, OWL, SPARQL, SWRL, CIDOC-CRM

No computer needed (pen & paper techniques)

Books: Seagaran et al: Programming the Semantic Web (O'Reilly)

Software: Protégé (OSS, Stanford) - http://protege.stanford.edu/

Examples from the Venice Project: Incanto System (1283-1453): Organization of commercial routes through an auction system (State owned). Source: Doris Stöckly, Le système de l'Incanto ... Brill 1995.
How to code that set of data ?
Simplest kind of dataset: Table (e.g. Excel), e.g.

Route	Departure	Destination	Departure_Date	Arrival_Date	Callling_at
R1	Venice	Alexandria	2/7/1422	5/8/1422	Corfu (2d), Candia (3d)
R2	Venice	Tana	23/6/1422	12/9/1422	Corfu (3d), Constantinople (4d)
R3	Venice	Bruges	12/3/1422	12/6/1422	Corfu (1d), Mallorca (3d) , Lisboa (3d)

With data in a spreadsheet, you can sort, display, print, edit

There are semantics in data table: Name of the columns
Limits : no (complex) querying, limited number of columns

-> Relational Database give you these options: multiple tables which you can join

Route
Place
ID
DepartureID ----------> ID
DestinationID GPS coord
Departure date Stop
Arrival date
Stop1 ID ----------> ID
Stop2 ID Length of stay
Stop3 ID

Problem: As data grows and extends, Schema migration: painful process

Simple Excel like tools don't have capabilities for complex querying and also have limit on information that can be stored, so the way out is relational databases. These too become problematic when data grows and these have to be mmigrated along with schema. So using metadata can be a way out which is future proof as well. you can keep adding more data to exisiting one as and when there happens to be by improving the metadata.

Make schema generic (EAV: http://en.wikipedia.org/wiki/Entity%E2%80%93attribute%E2%80%93value_model)

Route

ID ->
Name

Logical encoding
Subject Predicate Object: RDF Statement
(R1 Departure Venice)
(R1 Departure-date 2/7/1422)
this can also be presented as a graph

R1 -> departure -> Venice <- departure - R2
|
departure-date
|
v
2/7/1422
Syntax of RDF: subject, predicate, object
Data presented in RDF can also be presented using graphs with subject and object forming the nodes (known as **resources**) and predicate as edge between the two.

In the RDF (Resource Description Framework) Model, nodes of graphs are called Resources
RDF was defined with the idea of making data sharable across teams, so each resource is given a: URI (Unique Resource Identfifiers) permits to avoid ambiguities
Best known form of URIs are URLs (Addresses of web pages). URIs extend this concept to any unique identifier of a resource (they don't need be related to a concrete Web page, e.g. isbn: 987123456 for a book)
In RDF: predicates are always resources. The URI for a resource represented in RDF is the URI Ref (e.g. http://www.w3.org/1999/02/22-rdf-syntax-ns#type) It is good practice that there is a corresponding site, if you type this into the Browser, but not a requirement
With namespaces, you can shorten long URIs as the one above to rdf:type
Serialized representations of RDF-data can become complicated. There are multiple serialization formats: e.g. RDF-XML, N3-Triples, RDFa (RDF in attributes e.g. of HTML-pages)
A set of URIRefs is known as a Vocabulary, e.g. for our maritine route examples

What SQL is as query-language for relational database, SPARQL is the language to query RDF-graphs:

- Departure (?x1, Venice) //find x1 which departures from Venice
- Captain (?x1, ?x2) && Gender(?x2, Women) //such that x2 is captain of x1 and x2 is of gender Women

(all the routes to Venice that have a Women Captain)

With RDF coding, we can also write rules to infer new triples

- hasParent (?x1, ?x2) ^ hasBrother (?x2, ?x3) => hasUncle (?x1, ?x3) // x1 has x2 as parent and x2 has x3 as brother that implies (means) x1 has x3 as uncle

This is also a way of detecting possible incoherence in the set of knowledge coded in the triple store -> SWRL (Sematic Web Rule Language) is one standard language to do this.

ONTOLOGY framing :

Once you think you know that something exists, the next step is to put it into a relation to other objects. We don't do this for every possible thing, just for the entities in our problem domain.

- (Venice is_a Place)
- (Corfu is_a Place)
- (Place has_lat latitude)
- (Place has_long longitude)
-

Ontology can be represented as RDF and stored as graph alongside RDF.
OWL (Web Ontology Language)

Terminology statements

- ex:Bridge rdf:type rdfs:class
- ex:Bridge rdfs:subclass ex:Place

Assertion statements

- ex:Rialto rdf:type ex:Bridge
- ex:RialtoCons ex:broughtIntoExistence ex:Rialto

It is expressed in the same language. Eg "Rialto" is a "Bridge".

Ontologies themselves are resources that can be associated with an URIref. It is relatively easy to create your own ontology using a software like **Protégé**. But some ontologies aim at being "Universal": **CIDOC-CRM** is an ontology for Cultural heritage (about 20 years of work, ISO 21127, 100+ schema, very stable)
https://en.wikipedia.org/wiki/ISO_21127

In CIDOC-CRM, the modelling is event-centric: The underlying idea is to model change, not state: Instead of coding the birthdate of an actor, it is better to code the event of its birth:

- not: (Person is_born 1900-1-1)
- but: (Person hasBirthEvent BirthEventP1)
- (BirthEventP1 hasDate 1900-1-1)

Actors are thought to related to physical items only through temporal events.

CIDOC-CRM has dozens of predefined Properties (had participants, carried out, ...)

CIDOC-CRM has also a very interesting model for places. What is hard about places? What is the coordinate of a city? Things can happen on mobile places (like a boat). The question "where is it" can either specify a position (coordinate) or a physical thing (the ship, at home, where you might not know where "home" is) -> E53 Place can only be defined relative to a larger object (Cities are moving over time)

Influence is very hard to model (Actor A1 has taken notice of Object O2).

Problem in the RDF-Model: Qualification of facts (who stated it, how certain is it).
Technical trick:

- (RialtoReconstruction hasTimeSpan 1588-1591)

-> Reifed RDF

- (s1 rdf:subject RialtoReconstruction)
- (s1 rdf:predicate hasTimeSpan)
- (s1 rdf:object 1588-1591)
- (s1 metardf:reliability 0.8)
- (s1 metardf:source FKaplan)

Tutorial/Workshop Session 3: Pascal Föhr (@PFHist) (Basel University): Historical Sources Criticism in the Digital Age

Workshop stars with an online research exercice: "History of the Internet"
- approach
- tools
- results (kind of sources)
- usage

We're back in 15 minutes ;-) 15 minute-discussion by group is up!

How to get information?
Group 1: started with discussion, looking for a definition of the Internet. Definitional approach (OED), looking in bibliography of Wikipedia to link to NSF.gov website, Google search yielded internetsociety.org, ...
Group 2 used also Google. Consulted also a virtual catalog and used also Google picture search
Group 3: History of the Internet from unknown authorities, pictures, jstor general articles (and googlescholar), short histories of the internet -- a lot about specific aspects of internet rather than
Group 4: approach similar to other groups; concern with objectives (article? academic article? presentation? audience? credible authorities under these parameters)

- (Merci à xaentenza pour la référence: http://web.archive.org/web/)

Remarks about research:
- we talk about digital objects, not only texts.

- Tools: most groups used google at first
- **problems with Google**: "traduction" -- to quote in google is to quote in the world?
- Problems with google: you will always find something!
- - he tricked us!
- - assumed knowledge about what's in archives, catalogues, etc.
- - alternatives to google (https://duckduckgo.com) -> not to be confused with duck duck goose
- - "google is the end of the search engine as we know it"

What is a digital object?
- 01 - binary (code) ... "binary solo.... 0001110001000011001" <-- this makes 57881 :-) --> http://www.youtube.com/watch?v=CTjolEUj00g
- additional information
- file
- "machine readable"
- software
- immaterial (débat: you need for example a storage, you need also an interface)
- clarification: what is an object? (it's anything digital)

- Discussion:
- Photo example: http://www.worldpressphoto.org/awards/2013/spot-news/paul-hansen
- - image manipulated digitally - is it still a photo? is it manipulated? or is authentic?

- - concern with intentional manipulation
- - how many originals are there?
- - what makes digital changes special when we see the same techniques in analog photography
- - note: nothing has been "manipulated" in the photo -- there have "only" been filters added
- comparison to Trotsky being photoshopped out: http://www.ww2incolor.com/d/515471-4/Trotsky+out

Catharina Schreiber: "A digital source ('born digital source') is genuinely digital, multimodal and -dimensional, modifiable, process orientated as well as independent in time and space.
*multidimensional -
*modifiable - process changes

Problems working with and using digital objects

Accessibility/traceability: do we have access to the source, and how long do we have access to the source?
- manipulation of sources

Digital sources can be:
- not available
- updated
- manipulated

Authenticity/integrity

Saving/storing of digital information.
Is there a protection against saving? (have to save the whole page).
Example: sounds and music in http://www.incredibox.com/en/play#
- can we save it/ can we use it as a source. Problems of copyright. Can we save it from manipulation.
- basically the issue is the intangibility of digital sources (i think)
- tools with which we look at images

Presentation

Preservation
- original (and originality) lost

Destruction
- by accident or deliberately

Multimedia
- for literature or music, important to see versions

Publication services
- slideshare, RSS feed, etc.
- google maps
- accessiblity changes etc.

Manipulation
...

Internal/external criticism:
Internal criticism: problems if we use digital objects? (il est allé trop vite là...)

Possible solutions
1) archiving websites
2) digital footprinting
3) Zotero (saving the website as a pdf)
4) checking caches
Comment: we have citation methods that take into account mutability of websites, although students can't be trusted (to cite properly)...
In other words we need to trust the citation; the author needs authority.

Digital manuscript: it can be changed. If you can fix it, fix it. [I disagree! That's nothing new... the same problems existed with printed books]
Historical method
Versioning: know what you're using ("version 1.7" of the document for example)
Hash code: large numbers and digits. What you cand o with it you can download your object, put it through your program, and detect a manipulation but you don't know which one.
Metadata
Local data container
scientific platforms must implement security features
Researcher driven archiving: Institution with governmental support
Bunker archiving - network systems (one of); put it in a cave and never touch it again. [aka hoarding] Accessibility issues.

Suggestions for scientific writing
- figshare: http://figshare.com

Questions
a) Difference between the digital documents and print documents not so big. Ex: digital manuscrits, but a printed book... also has many editions (ex. 19th century and list of errata) Internal criticism (errata, remarks) vs external criticism (several libraries, so if you want to change something, you have to go to all libraries and change that thing you want to change).
b) Every question asked is equally applicable to printed matter (consensus in the room? seems so)
-- suggestion about changing terminology
pascal.foehr@unibas.ch

Tutorial/Workshop Session 3: Guido Koller (@GuidoKoller) (Swiss Federal Archives): Interpretation of Digital Records: The Swiss Federal Archives' case

According to Koller, problems relating to archiving/preservation and metadata management are at least partly solved. Still at stake are however questions on how to make records best available and how to interpret digital records.
Archives have a 30 year ceiling, newer documents are available for consultation on request
Looking at the archive, how it works with digital records
He discusses search problem with OCR (e.g. numeric data/statistics that are printed as text strings)

short intermezzo: DODIS search (http://db.dodis.ch)

Archives in the digital age (BAR perspective)
1980s: digital born data created in the administration
- oldest data: statistical office; magnetic tapes are lost (hardware not available, magnetic charge lost)
1990s: digital finding aids
2000s: digitalization, archiving digital data
- OAIS standard
2010s: big issue: context information
- until now essential distinction between primary data and metadata
- fulltext searchching makes metadata less important; for contextualisation however metadata matters
- (relational) databases are flattened (SIARD tool)

User needs
- data driven analysis; PDF is not very suitable for data analysis
- worst to best quality: Pdf->Excel->CSV (Comma-separated value)->XML- >Permanent Link/linked data
- collaborative web (e.g. user generated comments)

Conclusions/challenges
- developing infrastructure for access
- organizing spaces of information ("proto VRE")
- building up knowledge (internally to meet the challenges, externally as a data contributor)

Discussion
- metadata description on file or document level? consensus: the level is less important than the quality
- federated search on several archives/single point of access desirable
- abandon absurd rules (3 pictures of manuscript material per day and user in one cantonal archive)

Tutorial/Workshop Session 3: Martin Grandjean (Lausanne University)(@GrandjeanMartin): Introduction to Network Visualisation with GEPHI

Network Visualisation Tool: https://gephi.org/

Introduction to Network Visualisation
Import csv files (we don't talk about the preparation of the data today).

File for Nodes (points)// File for Edges (connexions between points)
Edges have a source and a target
There are several centrality measures: By degree (number of connexions) / By closeness (closeness to the entire of the network) / by Betweenness (bridge part of the network) / Eigenvector (pageRank-style centrality - the more well-connected nodes are linked to you, the more you are central). And the more you are central, the more your neighbors are too. [a really good piece of work on centrality indices http://link.springer.com/chapter/10.1007%2F978-3-540-31955-9_3]

RESOURCES:
Examples shown during the workshop:

- Twitter during a #DH conference : http://www.martingrandjean.ch/colloque-dhiha5-les-digital-humanities-se-deploient-sur-twitter/
- Archives mapping : http://www.martingrandjean.ch/analyse-de-reseau-nouvel-outil-exploration-fonds-archives/
- OpenData on State finances http://pegasusdata.com/2012/11/25/opendata-copinage-au-gouvernement-quand-lanalyse-de-reseau-vient-en-aide-au-journalisme-dinvestigation/
- GeoLayout plugin about european cities : http://www.martingrandjean.ch/archives-cartographique-geopolitique-megalo-urbanisation/

The **complete tutorial** to be able to analyse your facebook network : http://www.martingrandjean.ch/analyser-graphiquement-reseau-facebook/ Which is also a good **tutorial to GEPHI** if you avoid the part 1 about facebook, replaced by your own data or by the datasets prepared for the tutorial (data here: http://www.martingrandjean/gephi).

SATURDAY Unconference : http://www.dhsummerschool.ch/?p=262

F-123 DH team and the division of labour

A team of 10-12 people :

- - interdisciplinary
- - a big project composed of various projects and various methods
- - only 2 developers !
- - how to coordinate ? how to communicate between different backgrouds ?

- - making projects with different students coming form humanities & computer science : encountering some problems of communication and also of 'protecting' their own skills, professions, and so on
- - Fear to share knowledge, because they fear to loose the source of money (selling courses, working)
- - everyone have to know their place, speciality AND to communicate about it → to lower the fears | the roles have to be described at the begining of the project, documented, publicly (for the project participants)
- - need to buy "management for dummies" : project management skills

Coming from R&D Media / Publishing

- - Expertise + clear division
- - How to manage specialties ?
-

How to start a team ? How to get yourself recognize to get what you need.

- - In India they are so uge, and no central coordination...
- - experience : in the 90's to begin a project, they tried to build a social network, nowadays they are tools for that : scientific social networks, to have a lists of intersted people, then : the physical encounter is IMPORTANT.
-

How to balance between the need of freedom and having a common goals.

-

Solutions :

- - project management skills : humanities doesn't function like private corporation | agile methods, because of the need of the redefintion of the project, what happens often
- - clear definitions at the begining, making the communication easier
- - feat to share knowledge : sometimes technicians are underevaluated, just support. Be clear about the credits, including not only the leader, the students, and the technicians. How to _pay_ everyone, not only the money, but also name
- How to motivate the people to go in the same direction. Intersting in being on the paper, or only to code ? Need a clear definition what is expected of whom ?

- Clear strategies of recruitement. Team building, jobs profile.
- - the project manager need differents skills : in humanities, in management and in techniques, to be able to estimate correctly what ressources it takes to achieve a tasks, and to understand how, and why. These skills is not coming from a formation but from experience. Having the minimal understanding of the different disciplines to be able to communicate.
- - Making a team with students from different disciplines : to be able to realize difficult tasks, and this way the students can learn from the other students
- - if you know everything you don't need a team... (ironic) : that's why teams are good ideas
- - in a project, making little workshop to make different skilled people to learn to understand each other.

The projects are not exactly what the research processes are. For "projects", you need to have a common goal, and that everyone is willing to achieve it.

A necessity is to share information into the (future) DH community at large.

How to recruit people? How to find tips and solutions? Useful links:
- Website for questions and answers : the digital humanities q&a.
http://digitalhumanities.org/answers/
- Mailing list in dh. http://dhhumanist.org/
- DH commons: http://dhcommons.org/projects
- http://www.tei-c.org/Activities/SIG/Manuscript/
- French discussion list: https://groupes.renater.fr/sympa/info/dh
The picture of the black board
https://twitter.com/railyards/status/350898147767234560/photo/1 another one :
https://twitter.com/pauldoshea/status/350898215421349888/photo/1

F121 Visualization

The Swiss Society of History&Informatic proposes a call for paper in visualization : http://blog.ahc-ch.ch/?p=1128 (German) http://blog.ahc-ch.ch/?p=1133&lang=fr (French)
Questionning the order of paragraphs in the call, why to put visualization before methodology ?

Visualization: what's the difference between "technic" and "method" ? Scholars are not supposed to talk about technic, but in fact if we don't masterize the technic, it's not possible to elaborate methods.

distinction: visualization as a means of exploration -- visualization as a means of communication -- blended: visualization as a means to allow exploration (visual online application)

exploration: the possibility to dive into the data (--> qualitative analysis) right from the visualization is very desirable

(good) visualizations allow engagement with the public, create publicity

Resources

Example: http://totenbuch.awk.nrw.de/uebersicht/spruecheat (explorative)
how dance might be visualized: http://synchronousobjects.osu.edu/ (visualization as a means of communication)
actor analysis: http://www.martingrandjean.ch/network/ (using a gephi plugin)
sample gallery of d3.js: https://github.com/mbostock/d3/wiki/Gallery

F 021 To code or not to code
to code or not to code -

is xml, etc. also coding? is scripting?
how to use/adapt tools or -depending on the research questions- use existing tools in a unintendent way (p.ex: use hidden output of access and use it in a xml-table), thus hacking.
agree, we need to understand the tools and the code. for coding, there are other experts. p.ex. if websites are the source, you need to understand html to check for hidden information.

time and budgets limit, what is reasonably feasible
important to know, who to ask. how to make programmers interested? perhaps find people who studied both fields - IT and humanities.
need to introduce 'coding' in basic classes at universities. everybody has computers nowadays.
what to learn/teach? many programming languages, but need to understand the principles.
however: those in charge are often the old guard - who panic when thinking about computers.
question of generation? estonia: introduced coding to 7yrs old.
perhaps we need intermediaters: humanist have research question. than dh-center to ask, if tools already exist. and if not: is it feasible and how to translate this to coders
introduction to programming: change the way of thinking. but tools are not 'neutral'. need to find the tools which fit my research. be a broker
technical and humanist world are _very_ much separated. often, in projects both sides do not share a common language. digital humanists as 'team leader'/intermediate who understands both sides?
so, is there a 'real' scientific field of research for dh? is there enough science to pursue an academic career or any career at all at an university? is dh just a dead end?
should coding be regarded as a publication (today in austria, not at all - is invisible, more like doing nothing)

F-121 DH curriculum

Questions:

- How to teach DH ?
- What course would we have to attend during our studies to be good DHers today ?
- Bridging the gap : what content in courses?
- DHes as brokers, coordinators between disciplinary experts
- Should we teach DH or should we teach "digital" inside the disciplines ?
- Or should we teach DH or teach "humanities " for computer scientists, information scientists... ?

Beware : computer scientists are problem solvers, not "just coders"!
Interesting : Among participants, how many have taken a programme in DH ?

On a conceptual level, we should focus on methodologies, ways of thinking.. Find a way of thinking "technology*
Teach a way of thinking / mentality (methodology, modelling), not the use of specific tools. Also focus on how technology impacts your research practices.
Digital humanities will disappear when it will be comoin to use the current tools. When every humanistis feels confortable with new tech / news tools, what will DH be?
Who are our ideal co-workers? Experts in one (and just one) domain? Experts in more than one domains? Both models can work; different project cultures coexist. Another dependncy: availability of experts (IT staff that is passionate enough about e.g. text modification to do it at moderate salaries is scarce). A developer is not a digital humanist, he is a technician. A digital humanist would be a kind of translator between IT and humanists. Computer science is about certainty (yes/no), whereas humanities are about incertainaty (eg. "Yes, maybe in this particular context it is positive at 70-80%").

Expectations for a crash course of programming for humanists :
- Learn to work with texts (this would be a course for people working with texts)
- This is not necessary to separate disciplines

slides of a DH 2012 talk on a DH reference curriculum: http://de.slideshare.net/infoclio/prof-m-thaller-universitt-kln (contains list of current DH standards)

we need a digitatal humanities manual for dummies

l'umanista digitale è un ornitorinco!

what to learn? we need a guideline

F021 DH pedagogy

Lots of media and techniques, how are we supposed to deal with them ? The tools we "used" during these four days of summerschool:
-the lecturer/teacher
-the powerpoint (not always very well used)
-the framapad
-the chat of this framapad
-the twitter account
-our own notes (on a computer or on paper)

-the photographies we take of the slides
-the film made of the conferences
-informations on the website
-usb key with articles, informations etc.

The discussion turned around how to use at the best all these tools, are they all usefull, is it usefull to use them together...?

- On what support **preparing the course** // on what support to **give the lesson/**present the course // what to **give away to the students** ?

- - It's a big "waste" of time to have to prepare **several different supports.** But, as Claire Lemercier says it : it's maybe not such a waste of time because all students are different, and the different supports can "talk to" different students
- - the problem of the **red line** (fil d'Ariane) : not synchronized with your slides / Word preparation document. We are not sure that the students are sensitive to this
- -

- **"Powerpoint is evil"...**but asked for by the students, and sometimes compromises // http://www.wired.com/wired/archive/11.09/ppt2.html ; http://blogs.hbr.org/silverman/2010/04/powerpoint-is-evil-redux.html

- - bullets points for the structure (but not for the arguments = too salesman-like)
- - images /schemes to comment on
- - images to wake the students up
- - Tricks to "force" students to listen/look at the teacher : a black (or white) slide from time to time :)

- Thinking of **pads (Framapad**, piratepad, etc.) as a course support in the class : writing the structure of the course in advance, and the studentds completing/annotating

- There are some **handbooks for university teachers** UK, Canada "Preparing to teach (totally unthinkable in France to (i) exchange between the teachers, on their methods of teaching (supposed to be too personal) (ii) book on teaching for university teachers http://teaching.uchicago.edu/?/ctl-archive/course-design-tutorials/preparing-to-teach

http://www.google.ch/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&ved=0CCwQFjAA&url=http%3A%2F%2Fsanari-library.info%2Fwp-content%2Fuploads%2Fgravity_forms%2F1-ef40fdce7e01b3ab80b003b7af61a124%2F2012%2F12%2FPreparingToTeach.pdf&ei=_bbOUaWdH8G3O8qXgZgP&usg=AFQjCNHOqiNzeFCqmUyhuGuLszUyUFGXug&sig2=NlPS3DzugdHdqIg-ZgXZdA&bvm=bv.48572450,d.ZWU (=PDF)

- **Online courses ?** Seminars in place of /before/after the online follow-up ? Students seem not to like the online courses when it's purely online, and they are forced to (!!) (and marked

for) online discussions on different supports (chat, ...etc.), on University tools were the students are not "naturally" found (facebook, twitter ?), so they had to get used to it...

F-122 Critical humanities

cp. also
https://docs.google.com/document/d/1A4MJ05qS0WhNlLdlozFV3q3Sjc2kum5GQ4lhFoNKcYU/edit?pli=1

A copy of the latest version will be posted here for further editing and discussion.

QUESTION:

What would this pad become after #DHCH ?

Published in "Nature"? On a blog ? On a writeable platform ? On the dh summerschool website? http://lite.framapad.org/p/DHCH

Suggestion: We could split it into parts and link them to the Summer School programme on the summerschool website (after light editing ?) i like this idea :-) Me too!

Transformed into a words tags cloud ? :)
=> https://twitter.com/squintar/status/350587672546390016 Surely you are not talking about wordle :D

Change the setting, after creating, to change all to lower case. This will eliminate Dgital and digital as two different words.

We'll ask @yrochat to make this with R and compare with the Twitter wordcloud we'll produce after the DHCH (and a spaghetti monster, of course)

Actually, I'm planning to artificially add "sieste" and "Enrico Natale" as most frequent words. But chhhhhht. +1000

= A reference source for all the participants to keep track of what they have learnt and what tools/methods they have learnt/heard during DHCH and if they can use it for their research.

> Yes, but physically where ? not physically but may be virtually on the DHCH website. Well, an information is physically stocked on a server... Martin, stop.

> It would have to be edited, but then the website is a good option. OR someone thinks it's a visionary initiative and we try to do more (in this case any idea ? Nature or Science ?)

Can we keep the colors??? *Rainbow* :) +1

oh nooo, I've erased all the colors ;(=> can be found back in the history of the document

Become self-aware/skynet? and write the rest of the content by itself, in a digital and futuristic manner ? +1