DIGITAL SCIENTIFIC LITERATURE, BIBLIOGRAPHIES AND BIBLIOMETRICS

Enrico De Angelis[1], Tecla Caroli[1], Olga Beatrice Carcassi[1], chiara bernardini[1], qi.wang[2], Daniele Treccani[1], Marco Ricci[1], Simona Bravaglieri[1], liu.yuan[1], Ata Baniahmad[1], and Yuxin Yang[1]

[1]Politecnico di Milano
[2]Dipartimento ABC. Politecnico di Milano

March 11, 2019

Foreword

This text is about the course: *053643 - COMPLEMENTARY DOCTORAL SKILLS - Part 4: Digital (Scientific) Literature, Bibliographies and Bibliometrics* (2018/19 edition) and deals, mainly, with the work expected from its students.

The course aimed to introduce PhD Students to the use of Digital Scientific Literature, to train for systematic search within the many available information clusters, to perform quantitative analysis on their results and to organize their references in order to realize interesting literature reviews.

The ***first lesson*** introduces *The evolution of Scientific Literature (and of the scientist)*, from the first examples to the first twenty century's journals, then from the starting of the "great business" to the digital revolution, up to now (twenty-five years after the starting of Internet, everything is digital, the scientific publishing is still in a transition/evolution phase).

The ***second lesson*** gives suggestions, also at a very practical level, about *How to search (and to find) your scientific references* and how to (start to) realize a bibliography, starting from the many resources that Politecnico di Milano offers to its students and faculty members and the even more accessible for free, in the web.

RESOURCES	SERVICES	COURSES AND TUTORIALS	OPEN ACCESS	LOCATIONS

ELECTRONIC RESOURCES	CATALOGUES	THESES	ARCHIVES
E-JOURNALS	UNIVERSITY CATALOGUE	POST-GRADUATE THESES	HISTORICAL ARCHIVES
DATABASES	MATERIOTECA CATALOGUE	AND PHD	ARCHIVAL FONDS
E-BOOKS	MAPS	ONLINE THESES	ARCHIVAL INVENTORIES
ENCYCLOPAEDIAS AND	CADASTRAL MAPS	SUBMISSION AND	EXTERNAL INFORMATION
DICTIONARIES	POLITESI	ASSISTANCE	SYSTEMS
STANDARDS	EXTERNAL CATALOGUES		
WEB OF SCIENCE			
SCOPUS			

Figure 1: The menu of the Politecnico di Milano resources (see here).

onic In the ***third lesson***, we will see ***How to be found (searched/found/quoted/. . .)*** as researcher as

a basis to become a valued scientist, and we start reviewing the homework that will be used to assess th how much each student has reached the learning goal, that is to gain a deep experience in realizing a bibliography and to turn it in a literature review.

The **fourth and last lesson** is again a workshop, during which, the two authors of BIBLIOMETRIX (an R-based freeware), Massimo Aria and Corrado Cuccurullowill introduce their work to us, guiding us in its possible uses, with the bibliographic data collected by the students in the previous activities.

Classworks and Homeworks

First classwork

During the second lesson, five groups of two students were proposed and, for each group, some keywords of the two members were identified, with the aim to find a single common topic on which both the two students could work together. The joint proposed topics were the following ones:

1. Tecla Caroli and Olga Beatrice Carcassi - *Circular, R^n and biobased materials* LINK
2. Liu.Yuan - *Virtual learning application in medical education* *LINK*
3. Qi.Wang and Simona Bravaglieri - *Czech cubism , History and Architectural Design* LINK
4. Daniele Treccani and Marco Ricci - *BIM and GIS based solution in Built Environment management and retrofit* LINK
5. Yuxin Yang and Ata Baniahmad - *Electronic textiles, Development and Applications* LINK
6. Chiara Bernardini - *Climate change adaptation, Climate-related risk management*

Starting from these topics, the students extracted few keywords and experienced a bibliographic search in the SCOPUS, WOS, Google Scholar and zxccScholar databases.

First homework

The first assigned homework expects the students to perform the following work items:

- To **choose a topic**. The topic may be the same using during the first classwork or something else, based on other agreements with other students or even a personal one, more strictly related to their *research question* and start collecting bibliographic references.
- To perform a series of **bibliographic searches**, in one or more of the previous identified databases, and start collecting references and, if available, papers.
- To **take notes** of the rules results of the bibliographic searches and critically explain the evolution of the search rules, in the advanced search, as specified in a following paragraph.
- to extract, from the database, a series of interesting bibliographic items to be analyzed, classified and studied in the following activities, in an exchange format (BibTeX or others), also using softwares as Zotero, RefWorks, Endnote, Mendeley, Bibdesk or others.

Final work

The expected final work is a literature review. Well, may be it is too much for a 2,5 ECTS course. But we expect you to spend a good amount of time experimenting systematic bibliographic searches and the analysis that follows (and training yourself in that).

In the following chapters, there is some more about:

1. An approach for a systematic bibliographic searches
2. The State-of -the-Art
3. And the Literature Review

Bibliographic searches - an example

Suppose that our topic is mould induced decay (rotting) of wood based and natural, organic, construction materials. We choose the WOS database and execute the following searches, exploring the results and, for each one of the searches, trying to understand if alternative, similar key-words are more effective or not

We start from one of the main keyword, and we try the following searches (TS="topic", the number in black is the number of results):

(#1) TS ="building material": **3.501**
(#2) TS ="construction material": **3.032**
(#3) (TS="building material") AND (TS ="construction material"): **102**
(#4) (TS="building material") OR (TS ="construction material"): **6.431**

Since the intersection (3) between the set (2) and (1) of papers is very little, we check their consistency (verified) and we suggest to use their union (4).

Then we consider two alternative expressions "building product" and "construction product", as follows:

(#5) TS="building product": **235**
(#6) TS="construction product": **119**
(#7) TS="building product" AND TS="construction product": **3**
(#8) ((TS="building product") OR (TS="construction product")) AND ((TS ="building material") OR (TS ="construction material")) = **14**

Also the last two intersections are very little, so we propose the following expression:

(#9) (TS="building product") OR (TS="construction product") OR (TS ="building material") OR (TS ="construction material") = **6.768**

Then we go to the other field delimiting topic that may be identified with the following keywords: "rot(ting)", "mould", "fungi" and "moisture":

(#9) TS=rot: **38.486**
(#10) TS=rotting: **38.486**
(#11) TS="mould": **112.660**
(#12) TS="fungi": **198.518**
(#13) TS="moisture": **179.517**
(#14) (TS="rot") OR (TS="rotting") OR(TS="mold") OR (TS="fungi") OR (TS="moisture"): **498.885**

We can add few *wild characters* to check is the plural of the key-words, is included, automatically:

(#15) TS="rot*": **795.602** (that is really too much)
(#16) TS="rot?": **15.303** (that is really too low)

but the intrsection of #16 and #9 counts **1.646** records, so we took a look of them and we decided to avoid wild chars to "rot" and without "" because the set #9 NOT #10 is empty (**0**). We have similar results with

the term "moisture", while the case of the term "fungi" must be extended and include "fungus". "fungal". at the end, we choose the following set:

(#17) (TS=rot) OR (TS=mould?) OR (TS="fung*") OR (TS=moisture) OR (TS=myco*): **682.552**

and we change (adding plurals) to the following set:

(#9bis) TS="building product?") OR (TS="construction product?") OR (TS ="building material?") OR (TS ="construction material?"): **17.759**

that is a dreadful number of paper but, intersected with (#17) reports a more reasonable (but still high) number:

(#9bis) AND (#17): **1.796**

ALSO IF WE TRY TO REFINE IT, WE AREN'T ABLE to reduce significatively that number:

(#9bis) AND (#17) Refined by: [excluding] WEB OF SCIENCE CATEGORIES: (SOIL SCIENCE OR ENGINEERING MANUFACTURING OR MATERIALS SCIENCE CERAMICS OR RADIOLOGY NUCLEAR MEDICINE MEDICAL IMAGING OR MINING MINERAL PROCESSING OR ENERGY FUELS OR TRANSPORTATION SCIENCE TECHNOLOGY OR AGRICULTURE MULTIDIS-CIPLINARY OR ART OR GEOGRAPHY PHYSICAL OR GEOLOGY OR ME-CHANICS OR AGRICULTURAL ENGINEERING OR MANAGEMENT OR FOOD SCIENCE TECHNOLOGY OR PHYSICS NUCLEAR OR GEOSCIENCES MUL-TIDISCIPLINARY OR MATERIALS SCIENCE PAPER WOOD OR ENGINEER-ING AEROSPACE OR GEOCHEMISTRY GEOPHYSICS OR AGRONOMY OR PHYSICS PARTICLES FIELDS OR TELECOMMUNICATIONS OR MINERAL-OGY OR AUTOMATION CONTROL SYSTEMS OR BEHAVIORAL SCIENCES OR NUCLEAR SCIENCE TECHNOLOGY OR METALLURGY METALLURGI-CAL ENGINEERING OR MARINE FRESHWATER BIOLOGY OR COMPUTER SCIENCE SOFTWARE ENGINEERING OR COMPUTER SCIENCE THEORY METHODS OR ENGINEERING GEOLOGICAL OR ENGINEERING INDUS-TRIAL OR ENGINEERING OCEAN OR PHARMACOLOGY PHARMACY OR METEOROLOGY ATMOSPHERIC SCIENCES OR PHYSICS ATOMIC MOLEC-ULAR CHEMICAL OR HISTORY OR WATER RESOURCES): **1.330**

We looked and analyzed at all the retrieved papers and we clear those deal with soil (not soil based material), ceramics and metal molding

The State-of -the-Art and the Literature Review

The term "State-of-Art"

Everybody knows what a "State-of-Art" is or should be. We can, nonetheless, spend few words about it:

- The term "State of the art" is sometimes used as **positive attribute**. For example, you may consider "state-of-the-art" an equipment or a service that is the best (or among the best) equipment (device, technology, . . .) or service you may get, compared to . . . the state of the art. In this case, State-of-Art means "cutting edge" or "with the highest level of development", while it usually comes related to a specific sector, industry or scientific field and to a particular time.

- In Copyright and Patent practice, on the contrary, the term is used as **attribute with a non-positive meaning**. A "state-of-the-art" entity is a non-exceptional, non-innovative one that cannot claim any originality and to be an invention.
- The term "Art" may be used to identify the current (average or best) skills and the methods (i.e. the manufacture and craftsmanship techniques or technologies) in a specific field or activity so that "State-of-the-Art" means the **expected quality** of a product/service/activity without more precise specification, for example, in a written or implicit **contract**.

We will not use the term as an attribute. The "State-of-(the)-Art" is **a document** the describes the *Research* "State-of-(the)-Art" on one *Topic*, *Field* or wider *Area*.

A "modest proposal" in few paragraphs, about how to define a State-of-the-Art in three steps follows, with a warning. Never trust those who propose you a solution in three steps, for anything complex!
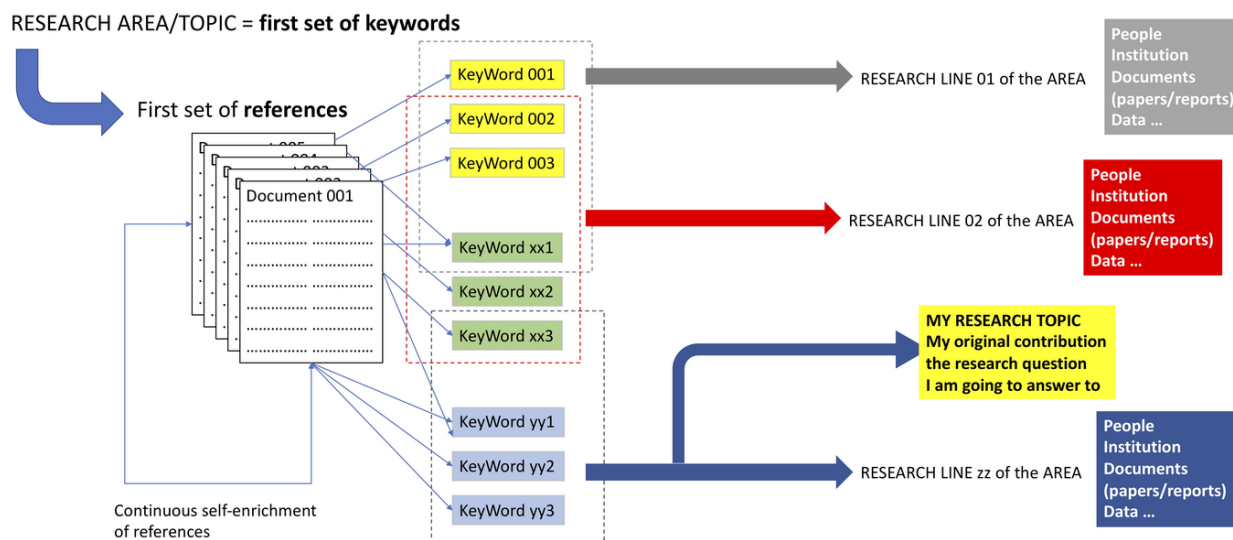
Figure 2: The "three steps method" for the definition of the *State of the Art*

A modest proposal - first step

You (as PhD Candidate) should have a **research area** or **field** clearly identified and to be able to define at least a first set of keywords. From this set of key-words you can perform your first search in one of the many abstracts and citations database we have already quoted. Just add the EU Research Projects databases and any other useful retrievable document collection you have access to.

The results of such a search will let you to collect a first set of references. Whatever the tool you will use to manage this list, you will have rapid access to their titles, authors and, at least, a summary for each one of them (books, papers, documents . . .).

Carefully **read** those few words, in order to understand if that item is or is not interesting for your study and filter it, eliminating any unsuitable content.

Interesting and relevant documents will be searched and found (I mean downloaded or acquired someway). Reading them, you will find other references quoted, enriching your collection in a never-ending updating process of your list.

Your Bibliography will grow, day after day, up to the end of your PhD path and beyond ... but you will soon reach a reasonably complete set of documents. Your first year bibliography will probably include already 80-90% of your final documents (if you are not going to change idea about your topic during your PhD, of course!).

A modest proposal - second step

Once you have a bibliography, you have to read, understand, assimilate and critically analyze each one of the documents you have collected (at least those that will become fundamental references for your work and your knowledge).

The best way to do this is to take notes and start writing: taking notes about their content and elaborating summaries that may become useful, providing some fundamental bricks of your future thesis (fundamental quotations, concepts ... however re-elaborated). Writing is always a good exercise, since the researcher's work is mainly reading (and critically understanding your readings) and writing.

You may even start writing structured texts and format them in a collection of information sheets about your readings, to be attached to your thesis, if it will include good pieces of work.

In any case, this series of information sheets will not be the main output of this second phase, but a more specific extraction of information.

Figure 3: You also collect your references to be sure that nobody else have already written your PhD Thesis, said J.Chaim, the author of PhD Comics, [see here, dated February 27, 2002].

The candidate should be able to extract, from each document (if realized, at the end of each information sheet), few keywords that will identify its main characteristics.

Nobody can substitute the PhD Candidates in doing this: it is a very personal activity. He (or she) can be helped, of course. But is the candidate who must identify the most significative characters of each document because its significance is strongly related to his/her specific research objective, questions, methodology, etc.

The classification of the references of a Thesis is also a very creative activity.

A modest proposal - third step

These keywords will take the Candidate very close to his/her state of the art, performing a synthesis action. Common keywords, among subsets, will immediately identify common characters among all the reference documents and, since these common characters will be mainly related to the objectives, the methodologies and the results obtained, in these documents, these keywords will, in practice, identify the main research lines that the candidate identify in the scientific and technical literature and among which he/she will position his/her research project, in a close future.

A modest proposal - in synthesis

Fully aware about the limits of each scheme, the proposed process is the following products and <*activities*>:

- <*identification of the research field*>
- first set of keywords
- <*abstract and citation search*>
- first set of references

- *<reference reading and writing>*
- information sheets
- second set of keyword
- *<keyword refinement>* and *<keyword aggregation>*
- research trends and lines

the list of the research trends and lines will constitute the index of the SoA.

The (PhD Candidate's) SoA

What is a SoA? The SoA that our PhD Candidates are expected to write, in their first 6 months work is, somehow the picture of the left zone in Figure 2, or the drawing of the context of the PhD Candidate's Research Question. It may be even a short paragraph, dedicated to each of the actual Research Lines intersecating with the Candidate's Research Question.

Not just a *set of lists* (the list of the people, of their Institutions, of the documents they produced in a fixed time-window, et cetera related to some specific keywords). It is a **story**. Although if a PhD Candidate is not expected to be a novelist, nevertheless he is expected to learn how to represent (in a graph, in a text *and* in a talk) where (s)he and his/her questions are positioned.

Whenever Candidates present theirselves must spend few words not only about their work and advancements but also about the State of the Art of their work.

From our point of view, a **SoA** is a text of 1000-2000 words. It will give a detailed picture of the most interesting *People* and *Institutions* **who** operate and operated on the *Research Topic*, their *Reasons* (**why**), their *Networks* and their periodic appointments (Conferences) and their *Activity* (**what** and **when** they have done, are doing and they project to do). The SoA will define (don't take literally this list):

1. The chosen Research Field and the "Problem Statement" (social/technical/economical/environmental) about which she/he will work.
2. The professional/social practice linked to that problem, its scientific basis, the knowledge about that research field.
3. An anticipation about the general objective of the Candidate's research work within that Problem Statement and how it will make the knowledge to advance.
4. The most important (and fundamental) people that have worked and are working in Candidate's Research Field
5. The research groups, the conferences and the meetings that have been organized, in the past, and are going to be organized in the future, in Candidate's Research Field;
6. The main research projects that have been funded, in the past, in particular EU Projects (see CORDIS), their results or attended results;
7. the main stakeholders (industrial, social, communities ...) directly involved or to be involved or just interested in solving the stated problem.

The biblography may be just an attachment but it is fundamental to continuously refer your statements to it.

State-of-the-Art or *Literature Review*?

What is the difference between a State-of-the-Art, as proposed in the previous paragraphs and a literature review? is a Literature Review simpler and smaller or longer and even more systematic? From our point of view a *Literature Review*, is a little different and something more.

We propose the following specifications for a Literature Review in five chapters:

1. The **first chapter** of your *Literature Review* will tell the WHOLE STORY of your search, explain the keywords and the syntax of your search, report the different databases you have used, and their results (what and where you tried the searches, what you obtained and how you changed your research strategy to optimize your results). This will, probably, be the last chapter you will write.

2. The **second chapter** will report the main information sources you have used and found. Once you have identified your first set of "interesting papers", you will be able to define the LIST of JOURNALS and other CONTENT REPOSITORIES that are more close to your research topic and questions. The list of journals may be organized like a bibliometric database and its content reported per PUBLISHER, with some bibliometric information like the number of papers identified in each journal or even the whole time distribution of the documents.

3. Researchers are social animals and you should be able to find the main meetings and other events during which a reasonably good number of researchers meet, in the world, to discuss their work. The **third chapter** will report the international or national CONFERENCES, their past proceedings, their past and the future ones where you would (should) present your work.

4. From the list of papers and conferences you will easily realize a LIST OF RESEARCHERS and INSTITUTIONS: some comments about their history will make the reading of this **fourth chapter** more exciting. The list may be even mapped in a world map.

5. RESEARCHERS write in JOURNALS and go to CONFERENCES. They are part of INSTITUTIONS and get funds from them. Often they get funds starting RESEARCH PROJECTS. If your subject is somehow related to a product or something similar, try to find PATENTS and a **fifth chapter** might enlist all them.

Also the Literature Review comes with a **Bibliography**, although it is a bit more than a simple annotated bibliography sorting a list references. A Literature Review is expected to identify:

- the main "schools of thought" or "approaches" or "trends"
- the main subtopics or competences (tools, methods, discoveries, . . .) you are referring to
- What has been written (significative, relevant) in the last n (5? 10? 50?? 200???) years about your specific topic or a general research question?

other resources

The author identify research items from different disciplines "political science", "social science", "urban studies", "art", "design" and others. The non-linear mesh of relationships intersects and the resulting mapping «visualises the landscape of the literature review across specific topics and disciplines, and demonstrates where they overlap». The mapping also reveals the gap in the knowledge.
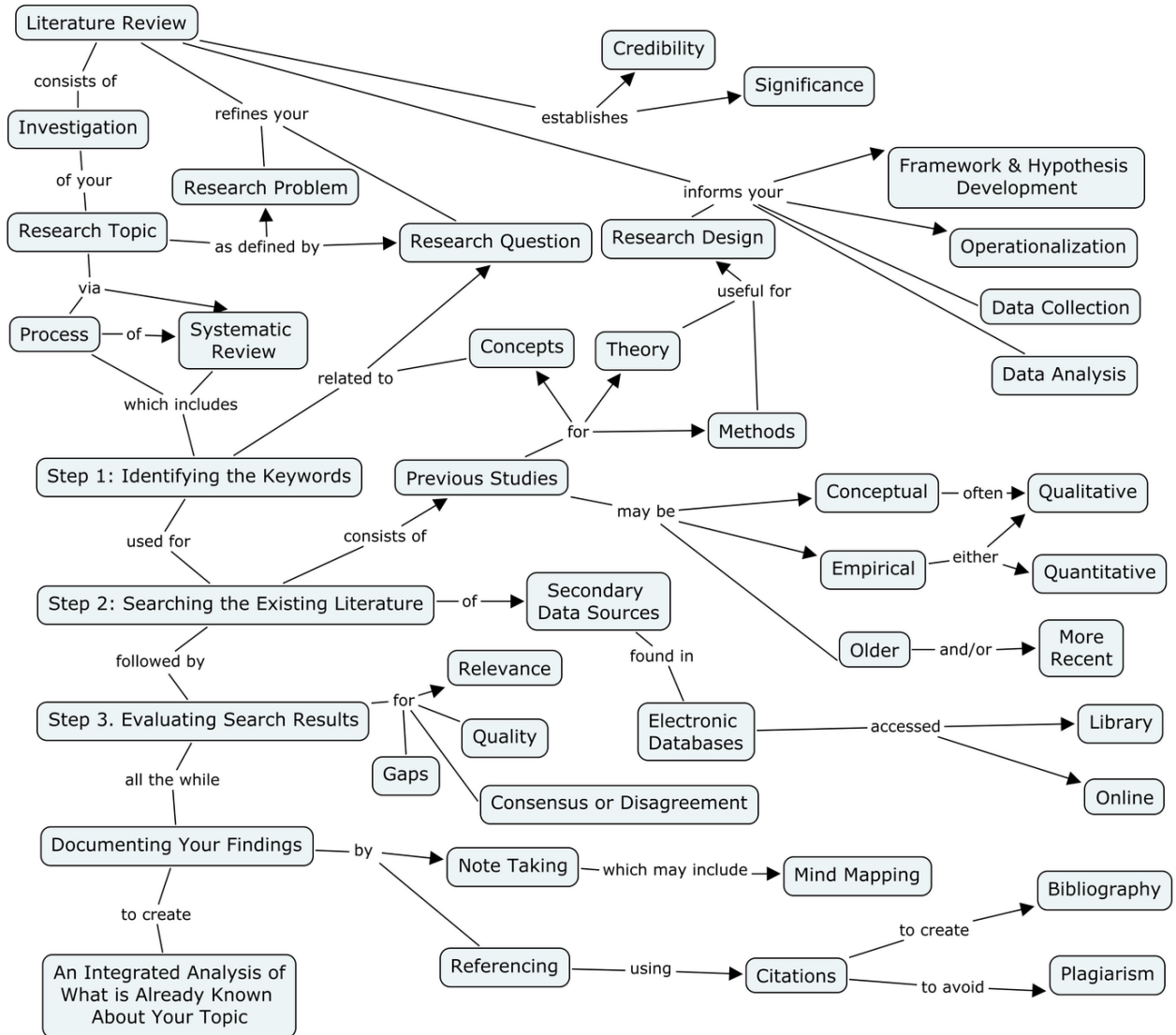
What is a literature review, why is it important and how do I do it?

Figure 4: C-map is an interesting mapping tool.

Literature Review
Literature Review on Lebanon

John Santrock

Ted Cantle
Del Felice & Wisler
Danish Youth Council
Home Office
Siobhan McEvoy-Levy
John Santrock
United Nations Development Programme
USaid
Youth Parliament (UK)

Smith, Jarman & Osborne

Marshall McLuhan

John Burton
Adam Kahane
Otto Scharmer

Tim Brown
Design Council (UK)
Tom Kelley
Jennifer Leonard
Triggs et al

National Social Marketing Centre (UK)
Bristol Social Marketing Centre

United Nations

Morton & Edwards
Sarah Pink

Bent Flyvbjerg
Steve Garner
Glaser & Strauss
Frederic Le Play

Gillian Rose
Edwards & Bhaumik
Patrick O. Marsh
Doreen Toutikian

Marcelle E. DuPraw
Ahmad Azem Hamad
John Paul Lederach

Community Cohesion Strategy (UK)
Social Cohesion Network (Canada)
Ted Cantle
Jane Jenson
Michael Toye

Ulric Neisser
Colin Robson
Thaler & Sustein

Jorge Frascara

M. El Machnouk
Carole Dagher
Kamal Salibi
Fawaz Traboulsi
Michael Young

Edward De Bono
David Moseley et al.
Bernard Pompëy
Robert Sternberg

Youth
Potential

Case Study
Methodology

IPA

Social Science
& Psychology
Methods

Social
Marketing

Conflict
Resolution
& Peacebuilding

Dialogue
& PSW

Visual
Ethno-
graphy

Media
Studies

Visual
Culture

Interdisciplinary
Methodology

Innovation

Visual
Communi-
cation

Social
Collaboration

Social
Cohesion

Sociology
& Political
Science
Theory

Cultural
Probes

Collaborative
Methodology

Service,
Industrial,
Architectural
and Interaction
Design

Social
Media

Communication
Design

Social
Integration

Design
Studies

Urban Studies

National Academies
Sharon Boggenpohl
Dorothy Deasy
Brenda Laurel
Bruce Mau
Tim Plowman
Desmond Rockfort
Patrick Svensson
Rosalind Sydie
Elizabeth Sanders
Dietmar Winkler
John Wood

Design
Activism

Communication
Theory

Social
Group
Dynamics

Social
Structures

Nicholas Abercrombie
AbdulRahman Annous
Emile Durkheim
Anthony Giddens
Stephen Hill
Samir Khalaf
David Lockwood
Patricia O'Hagan
Brigid Donelan Patricia
Lee Sigelman
Bryan Turner
Susan Welch
Jean Whyte

David Carson
Bauhaus
Jamie Reid
Stefan Sagmeister
Peter Saville
Jan Van Toorn
Andy Warhol

Sharon Jeannotte

Liz McQuiston

Frederick Boal
Salim Nasr
Zeina Saab

Jessica Dheere
Mark Zuckerberg

Big Society Network

Stephen Appold
Vincent Hong
Alexander Downes
Chaim Kaufmann's

Buchanan & Margolin

Erving Goffman
Emory Griffin
Richard Lanham
S.F. Scudder

Gaver, Dune & Pacenti

David Edmonds
Anthony Giddens
Safia Saadeh

Victor Papaneck

Peter Blau

Figure 5: "My first attemp at mapping the literature review in my PhD research". by Johanna Choukeir, here.