Subject metadata for humanities journal articles: Indexing consistency between a local repository and an external bibliographic database

DCMI 2019 panel *The Role of Metadata in Supporting Digital Humanities*

Ida Ahlström, Koraljka Golub, Joacim Hansson, Jukka Tyrkkö

Linnaeus University, Sweden
Purpose and aims

• Purpose
  • Understand potential impact of subject indexing on finding humanities research publications in digital platforms

• Aims
  • Compare subject index terms used for humanities journal articles
    • in a local university repository
      versus
    • Scopus
      versus
    • journal
Subject index terms: Repository

• Keywords
  • Entered by author at the time of deposit to the repository (optional field)
  • Additional keywords from the published article added by the librarian (if available)

• Categories
  • National Subject Category (mandatory field)
  • One or more categories chosen from a pre-set controlled list
  • Often corrected by the librarian reviewing the record (to at least a three-digit level)

• Research Subject
  • Matches University’s subjects, used for statistical purposes internally
  • The author enters one or more subjects from a controlled list
  • The librarian reviewing the record adds additional subjects if needed, to match the choices made in the field National Subject Category
  • (Optional field)
Repository interface: browsing

- Standard for Swedish categorization of research areas: “National subject category”
- Very broad
- These categories are used for official statistics on scientific publishing at the national level

- [URL to browsing]
Repository interface: simple searching

• Simple search
• Searches nearly all fields
Repository interface: advanced searching

- Keywords
- Research subject
- Category
Subject index terms:

- Extracted from a Scopus record
- Author keywords
- Controlled keywords
  - One field that combines values from various lists of controlled terms
- Uncontrolled keywords
Scopus interface: Simple search
Scopus interface: Advanced search

• Using field codes
• Humanities

Entering SUBJAREA(ARTS) will return documents that are classified under the subject area “Arts and Humanities”. The following subjects are classified under it:
- Arts and Humanities(all)
- Arts and Humanities (miscellaneous)
- History
- Language and Linguistics
- Archeology (arts and humanities)
- Classics
- Conservation
- History and Philosophy of Science
- Literature and Literary Theory
- Museology
- Music
- Philosophy
- Religious studies
- Visual Arts and Performing Arts
• **Author keywords:** keywords assigned to the document by the author
• **Index terms:** controlled vocabulary terms assigned to the document
• **Keywords:** a combined field that searches the other four keywords types

![Keywords Diagram]

- Author Keywords (AUTHKEY)
- Index Terms (INDEXTERMS)
- Keywords (KEY)
- Manufacturer (MANUFACTURER)
- Trade Name (TRADENAME)
The need and potential for an archaeology orientated towards the present

Holter, Cornelius
Linnaeus University, Faculty of Arts and Humanities, Department of Cultural Sciences.
ORCID ID: 0000-0002-0557-9651
2013 (English)
In: Archaeological Dialogues, ISSN 1380-2038, E-ISSN 1478-2294, Vol. 20, no 1, p. 12-18
Article in journal (Refereed) Published

Abstract [en]
The question 'Can an archaeologist be a public intellectual?' appears to express both an unfulfilled desire and a secret hope of an entire professional corps to count among them at least a few public intellectuals. I suggest that the state of the discipline of archaeology makes it harder, compared with other disciplines, for its professional representatives to address present-day issues and relate to public debates. I also suggest that maybe the most significant effect of the fact that society's public intellectuals generally do not have degrees in archaeology is that participants in public debates and policy makers are unaware of how various applications of archaeology and cultural heritage can benefit contemporary society. This potential will therefore have to be realized in different ways.

Place, publisher, year, edition, pages
2013. Vol. 20, no 1, p. 12-18

Keywords [en]
Public archaeology; archaeology in society; new heritage; applied archaeology

National Category
Archaeology

Research subject
Humanities, Archaeology

Identifiers
URN: urn:nbn:se:lnu:diva-26569
DOI: 10.1017/S1380203813000056
ISI: 000319987200003
Scopus ID: 2-s2.0-84878820344
OAI: oai:DiVA.org:lnu-26569
DIVA, Id: diva2:628837

Example in repository

• Keywords:
  • public archaeology
  • archaeology in society
  • new heritage
  • applied archaeology

• National category:
  • Archaeology

• Research subject:
  • Humanities, Archaeology
Example in Scopus

• Author keywords:
  • applied archaeology
  • archaeology in society
  • new heritage
  • public archaeology

• Indexed keywords:
  • archaeology
  • civil society
  • cultural heritage
  • policy making

Archaeological Dialogues
Volume 20, Issue 1, June 2013, Pages 12-18

The need and potential for an archaeology orientated towards the present (Review)

Holter, C.</sup>, Linnaeus University, Kalmar, Sweden

Abstract

The question 'Can an archaeologist be a public intellectual?' appears to express both an unfulfilled desire and a secret hope of an entire professional corps to count among them at least a few public intellectuals. I suggest that the state of the discipline of archaeology makes it harder, compared with other disciplines, for its professional representatives to address present-day issues and relate to public debates. I also suggest that maybe the most significant effect of the fact that society's public intellectuals generally do not have degrees in archaeology is that participants in public debates and policy makers are unaware of how various applications of archaeology and cultural heritage can benefit contemporary society. This potential will therefore have to be realized in different ways. © 2013 Cambridge University Press.

SciVal Topic Prominence

Topic: Archaeology | Heritage | Cultural resource
Prominence percentile: 81.179

Author keywords
• applied archaeology
• archaeology in society
• new heritage
• Public archaeology

Indexed keywords
• archaeology
• civil society
• cultural heritage
• policy making

GEOBASE Subject Index:
• archaeology
• civil society
• cultural heritage
• policy making
The need and potential for an archaeology orientated towards the present

Cornelius Holtorf
DOI: https://doi.org/10.1017/S1380203813000056 Published online by Cambridge University Press: 06 June 2013

Abstract
The question ‘Can an archaeologist be a public intellectual?’ appears to express both an unfulfilled desire and a secret hope of an entire professional corps to count among them at least a few public intellectuals. I suggest that the state of the discipline of archaeology makes it harder, compared with other disciplines, for its professional representatives to address present-day issues and relate to public debates. I also suggest that maybe the most significant effect of the fact that society’s public intellectuals generally do not have degrees in archaeology is that participants in public debates and policy makers are unaware of how various applications of archaeology and cultural heritage can benefit contemporary society. This potential will therefore have to be realized in different ways.

Keywords
Public archaeology
archaeology in society new heritage
applied archaeology
Analysis
Data collection

• Metadata of 649 articles in the humanities

  • Output of Linnaeus University
    • 2010-2018
    • Peer-reviewed journal articles

  • Found in repository http://lnu.diva-portal.org/

  • Date of extraction: 19 March 2019

  • English-language index terms only
Data subset (n=203)

- 29 articles with
  - Repository keywords
  - Repository category
  - Repository research subject
  - Scopus author keywords
  - *Scopus controlled keywords*
  - Keywords from journals

- 174 articles with
  - Repository keywords
  - Repository category
  - Repository research subject
  - Scopus author keywords
  - Keywords from journals

- The remainder do not have any Scopus keywords (68.7%, 446 of 649) or repository keywords (21%, 135 of 649)
  - In repository lacking probably because none found at the journal, not obligatory field
Qualitative analysis

• 29 articles with complete metadata
  • 5 have also uncontrolled keywords in Scopus
  • Main result: Scopus more subject access points through controlled keywords
    • “The need and potential for an archaeology orientated towards the present” (nr. 26)
  • Repository index terms
    • Keywords: Public archaeology; archaeology in society; new heritage; applied archaeology
    • Categories: Archaeology
    • Research subjects: Humanities, Archaeology
  • Scopus index terms:
    • Author keywords: Public archaeology; archaeology in society; new heritage; applied archaeology *same as repository keywords*
    • Controlled keywords: archaeology; civil society; cultural heritage; policy making
  • Journal keywords: Public archaeology; archaeology in society; new heritage; applied archaeology *same as repository keywords*
“At the threshold of retirement: From all-absorbing relations to self-actualization” (nr. 2)

- Repository index terms
  - Keywords: Cultural structures; emancipation; gender equality; I-we balance; retirement; Sweden
  - Categories: Cultural Studies; Sociology
  - Research subjects: Humanities, Cultural Sociology

- Scopus index terms
  - Author keywords: Cultural structures; emancipation; gender equality; I-we balance; retirement; Sweden same as repository keywords
  - Controlled keywords: gender; human; life; retirement; self actualization; social structure; Sweden; welfare; female; human relation; lifestyle; perception; psychology; retirement; satisfaction; self concept; Sweden; Female; Humans; Interpersonal Relations; Life Style; Personal Satisfaction; Retirement; Self Concept; Social Perception; Sweden; Women, Working

- Journal keywords: Cultural structures, emancipation, gender equality, I-we balance, retirement, Sweden same as repository keywords
• “Freedom to Divorce or Protection of marriage? : The Divorce Laws in Denmark, Norway and Sweden in the Early Twentieth Century” (nr. 3)

• Repository index terms
  • Keywords: Scandinavia; Scandinavian legislative cooperation; divorce; marriage; laws; women’s status; gender equality; individuality
  • Categories: Economic History
  • Research subjects: Humanities, History

• Scopus index terms
  • Author keywords: Scandinavia; Scandinavian legislative cooperation; Divorce; Marriage; Laws; Women’s status; Gender equality; Individuality same as repository keywords
  • Controlled keywords: article; Denmark; divorce; economics; education; ethnology; female; gender identity; history; individuality; jurisprudence; legal aspect; marriage; Norway; psychological aspect; spouse; Sweden; women’s rights; Denmark; Divorce; Gender Identity; History, 20th Century; Individuality; Judicial Role; Jurisprudence; Marriage; Norway; Spouses; Sweden; Women; Women’s Rights

• Journal keywords: Scandinavia, divorce, marriage, laws, Scandinavian legislative cooperation, women’s status, gender equality, individuality same as repository keywords
“Early Upper Paleolithic chronology in the Levant: new ABOx-SC accelerator mass spectrometry results from the Mughr el-Hamamah Site, Jordan” (nr. 4)

• Repository
  • Keywords: C-14 dating; Middle Upper Paleolithic transition; Jordan in this example entered by author since different from journal keywords
  • Categories: Archaeology
  • Research subjects: Humanities, Archaeology

• Scopus
  • Author keywords: ABOx-SC pretreatment; AMS; Biocultural evolution; Middle-Upper Paleolithic transition; Radiocarbon dating same as journal keywords
  • Controlled keywords: accelerator mass spectrometry; anthropology; archaeology; artifact; geochronology; human evolution; lithic fragment; Paleolithic; radiocarbon dating; charcoal; humic substance; analysis; chemistry; fossil; human; humic substance; Jordan; mass spectrometry; paleontology; procedures; radiometric dating

• Journal keywords: ABOx-SC pretreatment; AMS; Biocultural evolution; Middle-Upper Paleolithic transition; Radiocarbon dating same as Scopus author keywords
• “The 2004 Indian Ocean Tsunami Catastrophe, its Survivors, Job and the Universal Features of Suffering: A Theoretical Study”

• Repository
  • Keywords: Suffering; Tsunami; Holistic; Health care; Philosophical analysis
  • Categories: Religious Studies; Health Sciences
  • Research subjects: Health and Caring Sciences

• Scopus
  • Author keywords: Suffering; Tsunami; Holistic; Health care; Philosophical analysis same as repository keywords
  • Controlled keywords: cultural anthropology; disaster; human; Indian Ocean; interview; literature; mental stress; psychological model; psychology; religion; survivor; tsunami; Bible; Culture; Disasters; Humans; Indian Ocean; Interviews as Topic; Models, Psychological; Religion and Psychology; Stress, Psychological; Survivors; Tsamis

• Journal keywords: Suffering; Tsunami; Holistic; Health care; Philosophical analysis same as repository keywords
• “Participation and Enjoyment in Play with a Robot between Children with Cerebral Palsy who use AAC and their Peers” (nr. 9)

• Repository
  • Keywords: Augmentative and alternative communication; Cerebral palsy; Children; Complex communication needs; Peers; Robotics; Play; Participation; Enjoyment
  • Category: General Language Studies and Linguistics
  • Research subjects: Humanities, Swedish

• Scopus
  • Author keywords: Augmentative and alternative communication; Cerebral palsy; Children; Complex communication needs; Enjoyment; Participation; Peers; Play; Robotics same as repository keywords
  • Controlled keywords: Diseases; Gesture recognition; Human rehabilitation engineering; Robotics; adult cerebral palsy; child; communication aid; female; human; male; middle aged; peer group; preschool child; recreation; robotics; social participation; Adult; Cerebral Palsy; Child; Child, Preschool; Communication Aids for Disabled; Female; Humans; Male; Middle Aged; Peer Group; Play and Playthings; Robotics; Social Participation
  • Uncontrolled keywords: Augmentative-and-alternative communication; Cerebral palsy; Children; Complex communication needs; Enjoyment; Participation; Peers; Play

• Journal keywords: Augmentative and alternative communication, Cerebral palsy, Children, Complex communication needs, Peers, Robotics, Play, Participation, Enjoyment same as repository keywords
Quantitative analysis

• Based on the same 203 articles
• Using Cytoscape, free tool
Categories versus repository keywords

- Dominant:
  - Archaeology
  - History
  - General language studies and linguistics
  - Specific languages
  - Languages and literature

- Clusters of clusters
  - General language studies and linguistics
    + Specific languages
    + Languages and literature
  - Archaeology + History + Cultural Studies

- 5 outliers
  - Other Sciences not Elsewhere Specified
  - History of Technology / Ceramics
  - Public Health
  - Social Work
  - Social and Economic Geography

- Large image
Categories versus Scopus keywords

- **Dominant:**
  - Same as with repository keywords

- **Clusters of clusters**
  - Same as repository keywords

- **3 outliers**
  - Other Sciences not Elsewhere Specified
  - History of Technology / Ceramics
  - Public Health

- **Large image**
Articles versus *repository* keywords versus categories

- Blue = articles
- Red = keywords
- Light blue = categories
- Keyword “Sweden” – largest node

- Large image
• Articles nr 13 and nr 5 share the keyword “ethics”
  • Nr 13 is about tourist photography
    • “Gazes and Faces in Tourist Photography”
  • Nr 5 is about nursing and end-of-life care
    • “Futile cardiopulmonary resuscitation for the benefit of others: An ethical analysis”
Scopus author keywords versus repository keywords

• Similar

  • Average Jaccard coefficient = 0.75

  • Average number of keywords in repository but not in Scopus record per article = 1.1

  • Average number of keywords in Scopus but not in repository per article = 1.2
Concluding remarks
Summary...

• Most humanities articles in Scopus have no subject keyword access (68.7%)
• Most humanities articles in the repository do have subject keyword access (79%)

• Most humanities articles in both databases rely on article keywords rather than controlled
  • Only 4.5% (29 of 649) have controlled keywords in Scopus
  • Research subjects and National categories assigned to all articles (100%) in repository, but use is more for browsing than searching

• Both search interfaces
  • No disambiguation, narrower topic, related topic based on relationships in KOSs
  • Browse limited to a rather small number of categories
...Summary...

• Lack of consistent indexing policy in Scopus
  • Some controlled terms too broad like “article”, “analysis”, “human” → may lead to high recall at the expense of precision
  • Conceptual duplicates like “Scandinavia” by author, while “Denmark”, “Norway”, “Sweden” in controlled keywords; or “laws” by author versus “legal aspect” in controlled keywords
  • Word form (singular/plural) variations
  • Coordination variations (some more post-coordinate, others more pre-coordinate)

• No mappings between different KOS / no alignment across databases for keywords

→ This does not fulfill the well-established bibliographic objective to enable finding everything that there is on a certain topic
...Summary

• Compared to repository, Scopus provides many more access points if using controlled keywords:
  • In 4.5% of cases
  • In the 31% that have any Scopus keywords, they are the same as repository keywords in 86% (174 of 203) articles

• Within repository:
  • Categories and Research Subjects add mostly broader, disciplinary access points (as expected)
→ As with Scopus, lack of indexing policy for authors (no instructions, not obligatory, no KOS for keywords), also does not fulfill the established bibliographic objective
Future work

• Compare in more detail across KOSs

• Interviews with authors and users for analysis of implications on findability

• Interviews with database owners / indexers on policies

• Look into importing more subject keywords to the repository (even when authors added, from KOS...)

• Auto-suggestions at the time of deposit / social tagging