Search our Collections Queries Analysis Notes

A note on the aggregated data and data analysis

<u>**What did I do?**</u>

I grouped and coded search queries from a sample collected by the National Museums Scotland's staff from their Search our Collections (SoC) portal from 2020-2021:

- With frequencies from 90-1,909 (search queries with highest frequencies)
- With frequencies from 15-30 (search queries with lower frequencies and higher unique queries/even lower had too many unique queries for the timescale of this project)

I coded these queries based on the matrix by Klavans et al (2014)[1] project (see Figure on page 2) for defining level of detail/description of subject matter portrayed by images and applied this to user queries. I then coded queries for common general subject matter, and for strings of queries grouped together common modifiers/facets.

<u>**Tweaks to the matrix**</u>

- In category **S2** (individually named people, group, or thing) – named objects and individual people were <u>uncommon</u> in SoC queries, but I thought it would be useful to use this 'specifics/detailed' category to place species' names, technical terms, and culturally specific terms for objects, and very specific materials.
- In category **A1**, I grouped together spiritual objects and beings as they could be connected to more abstract ideas.
- In category **G4**, we grouped words for holidays, seasons, and general time periods (e.g., medieval) as uncommon in the Museums' collection metadata, placing only specific numeric dates/date ranges in S4.

<u>**Caveats**</u>

- The context is lost in the spreadsheet of queries, so I considered each line in the spreadsheet as an individual search even though they may have been combined: For example, object name, description, and collections were all separated in the spreadsheet but may have been combined in interesting and unique ways. The only words not separated, were those contained in the same line, for example, in 'description' someone may have put 'wine, vessel'. This means that the number of search queries in the spreadsheet and the sample could be a lot less.
- There are duplicates of object records, which may also skew the numbers and the analysis.

---

[1] Klavans, J. L., LaPlante, R., & Golbeck, J. (2014). Subject matter categorization of tags applied to digital images from art museums. *Journal of the Association for Information Science and Technology*, *65*(1), 3–12. https://doi.org/10.1002/asi.22950

Overview of data sample

Total unique queries for 2020-2021: 70,874

Total queries including repetitions for 2020-2021: 570,990

Total unique queries in sample: 1779

Total queries in sample (sum of all frequencies): 168,160 (Excluding empty queries) 168,970 (810 searches with nothing in query)

TABLE 2.   Subject categorization matrix (Armitage & Enser, 1997; Shatford, 1986).

		(G) Generic (Preiconography)	(S) Specific (Iconography)	(A) Abstract (Iconology)
1	Who?	G1: Kind of person or thing (*woman, dog, sunlight*)	S1: Individually named person, group, thing (*Mona Lisa, Daddy the pit bull, American*)	A1: Mythical or fictitious beings (*Clio the Greek muse, unicorns*)
2	What?	G2: Kind of event, action, condition (*football game, war, buying and selling, death*)	S2: Individually named event, action (*2009 Super Bowl, Battle of the Bulge, JFK's funeral*)	A2: Emotion or abstraction (*a woman buying fruit representing commerce or a funeral representing sorrow*)
3	Where?	G3: Kind of place: geographical, architectural (*city, jungle, inside*)	S3: Individually named geographical location (*New York, Mars*)	A3: Place symbolized (*Mount Olympus, paradise*)
4	When?	G4: Cyclical time: season, time of day (*spring, night*)	S4: Linear time: date or period (*June 1885, Renaissance*)	A4: Emotion, abstraction symbolized by time (*spring symbolizing youth or fertility*)
V	Visual elements	V: Shapes, forms, and colors (*square, line, red*)		
O	Other	O: Items that do not fit into the above categories		

JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY—January 2014   7
DOI: 10.1002/asi

Figure 1 Subject categorization matrix reworked by Klavans et al.,2014, p. 7

The following sections describe the categories of terms observed in the SoC spreadsheet and their subject matter by most common to least commonly searched.

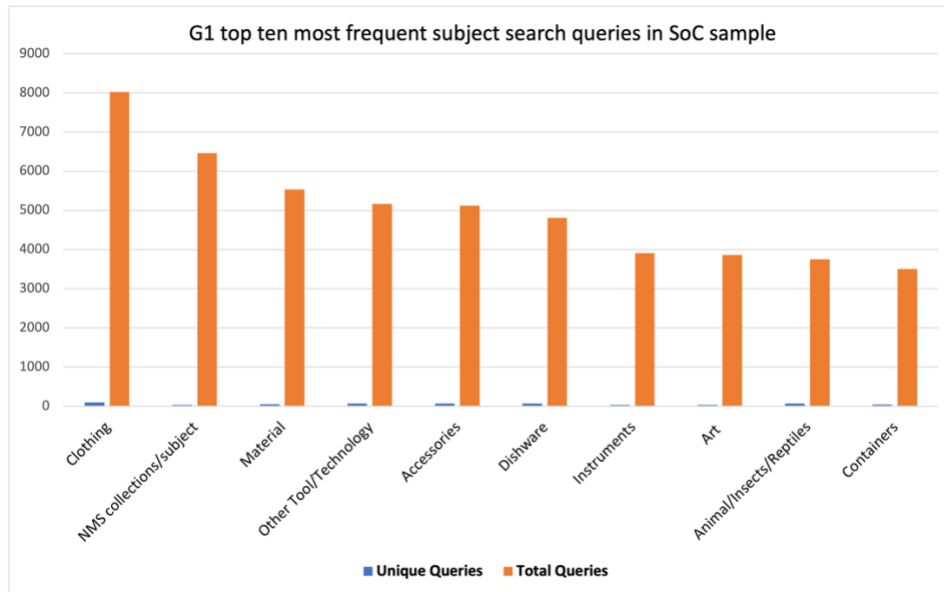## G1 General kind of person or thing

Unique searches: 885

Total searches including query frequencies: 76157 (including searches by collections/theme)

G1 Search Subject Queries	Unique Queries	Total Queries
1. Clothing	91	8021
2. NMS collections/subject	31	6460
3. Material	49	5539
4. Other Tool/Technology	69	5172
5. Accessories	66	5122
6. Dishware	65	4807
7. Instruments	29	3907
8. Art	28	3867

9. Animal/Insects/Reptiles	66	3750
10. Containers	42	3509
11. Weapons	43	3621
12. Sculpture/Model	33	3090
13. Communication	39	2735
14. Vehicles	35	2624
15. Astronomy/Navigation	13	1970
16. Furniture	18	1548
17. Design	19	1341
18. Death	19	1314
19. Health	19	1031
20. Identity	11	975
21. Features	13	638
22. Toy	8	589
23. Games	8	558
24. Awards	8	534
25. Minerals	8	503
26. Money	5	478
27. Light/lighting	8	236
28. Photography	5	399
29. Vehicle models	4	388
30. Music players	5	335
31. Pottery	3	265
32. Plants	4	244
33. Symbolic	2	207
34. Food/Drink	7	143
35. Body-part	4	78
36. Roles	3	63
37. Dinosaurs	2	37
38. Genders	2	36
39. Fossils	1	23
Total	**885**	**76157**

Facets/Narrower for strings:

1. Type of object (Arm ring, Foot plough, wine vessel)
2. General size (Miniature)
3. Gender (Woman, boy, man, baby)
4. Quantity (Group)
5. Part (Coffin base, Belt buckle)
6. Place (Japan/China/Chinese)
7. Aesthetic feature (Beadwork, embroidery, plaid)
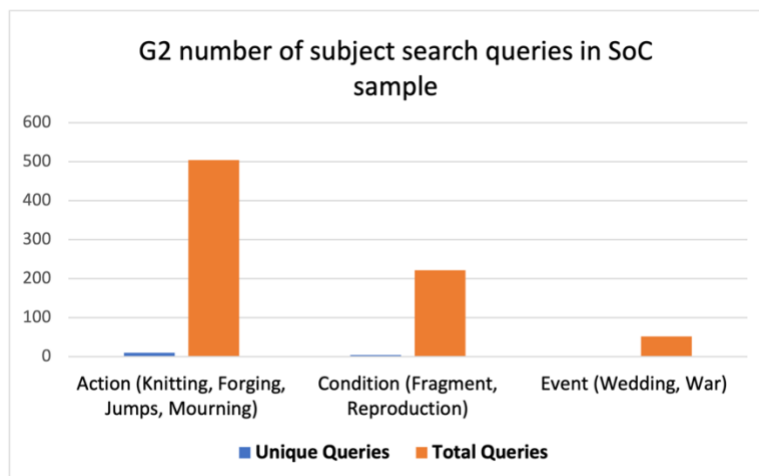8. For a particular person...
   a. Role: Actor, Chief etc.

G1 top ten most frequent subject search queries in SoC sample

G2 General What (Action/Condition/Event)

Unique searches: 16

Total searches including frequencies/repetitions: 778

G2 Top subject queries	Unique Queries	Total Queries
1. Action (Knitting, Forging, Jumps, Mourning)	10	504
2. Condition (Reproduction)	4	222
3. Event (Wedding, War)	2	52
Total	**16**	**778**

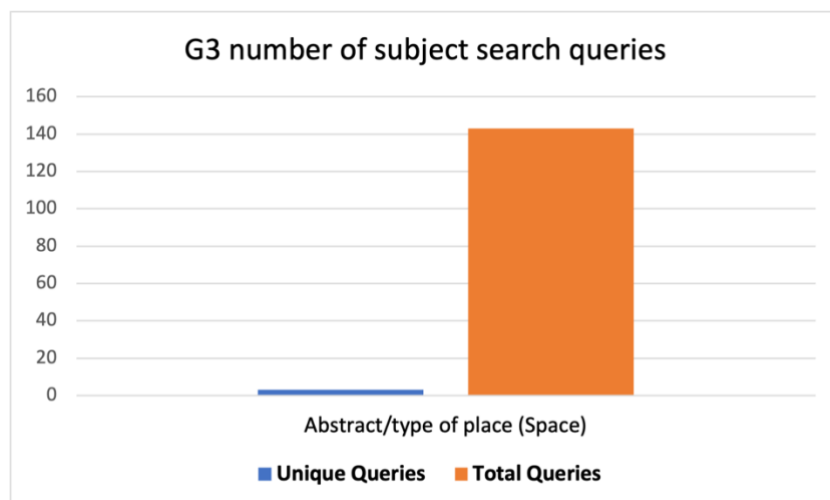G2 number of subject search queries in SoC sample

G3 General Where/Place

Unique searches: 3

Total searches including frequencies/repetitions: 143

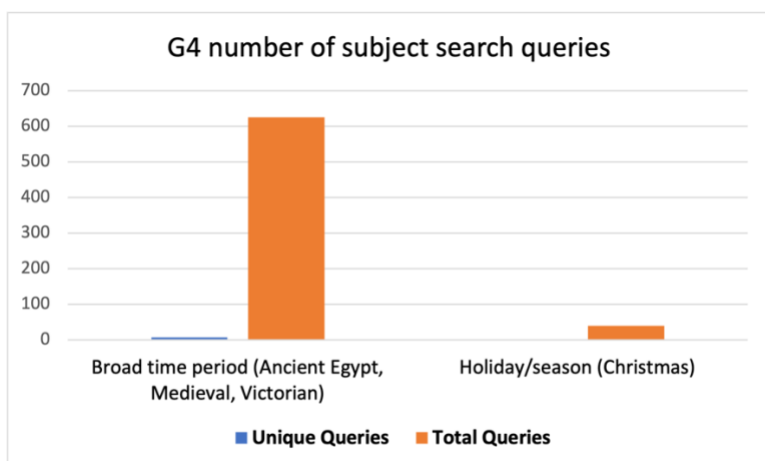| G3 Search Subject Queries	Unique Queries	Total Queries
Abstract/type of place (Space)	3	143
Total	**3**	**143**

G4 General When/Time

Unique searches: 9

Total searches including frequencies/repetitions: 666

G4 Search Subject Queries	Unique Queries	Total Queries
1. Broad time period (Ancient Egypt, Medieval, Victorian)	7	626
2. Holiday/season (Christmas)	2	40
Total	**9**	**666**

Visual (Colour, age, shape, size)

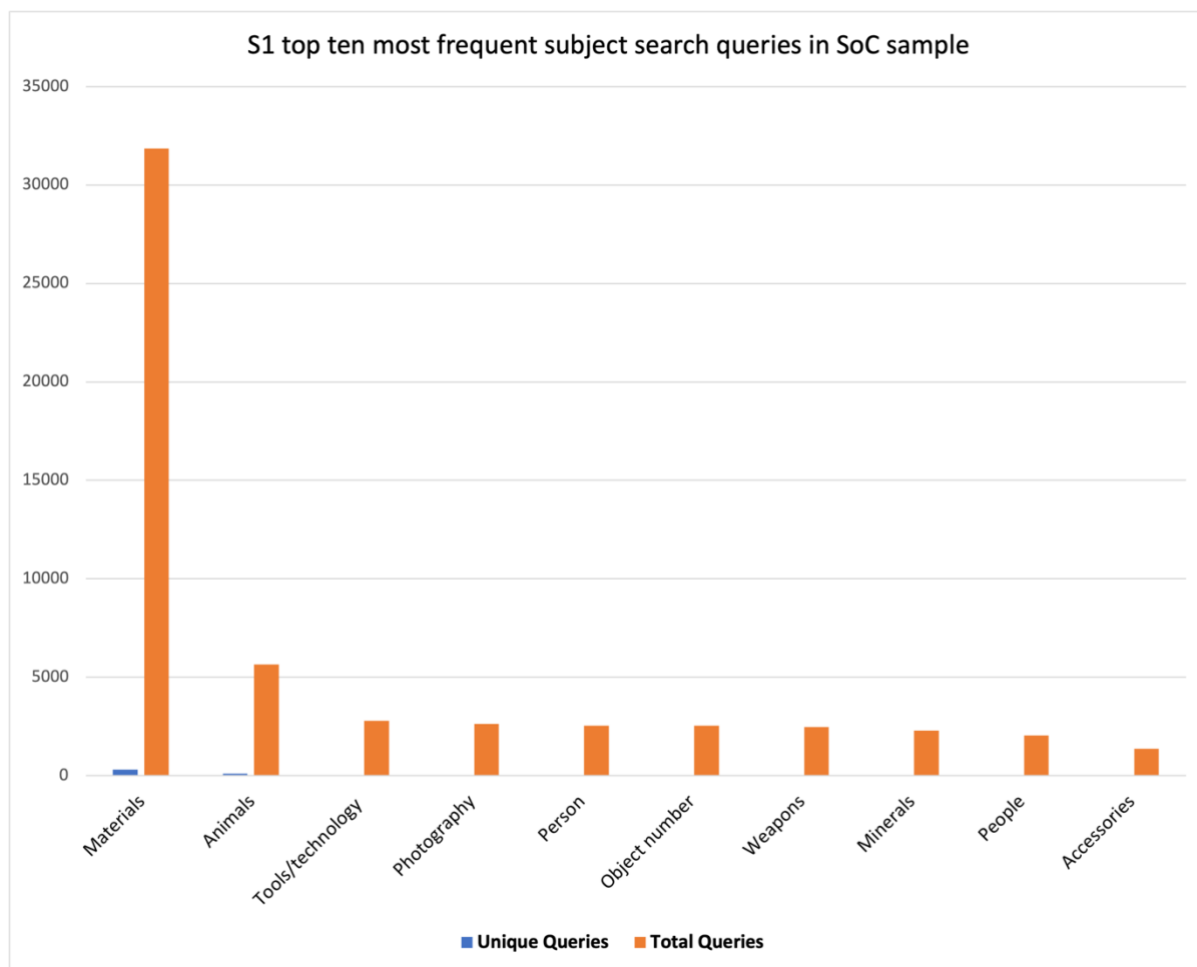| V Search Subject Queries	Unique Queries	Total Queries
Size, condition, age	5	258
Total	**5**	**258**

S1 Specific named person, people, thing (Included specific materials, species names, culturally specific or more difficult technical terminology)

Unique searches: 708

Total searches including frequencies/repetitions: 65 682

S1 Search Subject Queries	Unique Queries	Total Queries
1. Materials	301	31856
2. Animals	93	5654
3. Tools/technology	31	2780
4. Photography	11	2616

5. Person	40	2534
6. Object number	15	2528
7. Weapons	22	2474
8. Minerals	35	2284
9. People	27	2043
10. Accessories	15	1357
11. Art	10	1315
12. Dishware	18	1076
13. Instruments	12	1069
14. Clothing (Mantua, Abaya)	11	825
15. Vehicles/Transportation	7	709
16. Plans	7	703
17. Containers/cases (Tihu, amphora, shabti)	10	652
18. Games	2	535
19. Communication (Stele, Coptic)	7	528
20. Health	2	312
21. Plants	3	290
22. Money	4	283
23. Companies	4	247
24. Features	3	222
25. Vehicle models	3	211
26. Sculptures	1	147
27. Death	3	129
28. Music Players	2	128
29. Dinosaurs	2	56
30. Furniture	3	53
31. Pottery	3	50
32. Body-parts	1	16
Total	**708**	**65682**

S1 top ten most frequent subject search queries in SoC sample
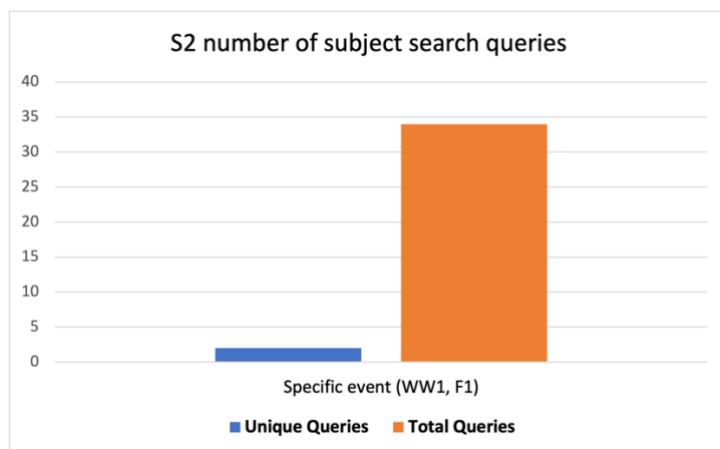
Facets/Narrower for strings

1. Part of... (Busk stay, Sarong kris tajam, scabbard sword)
2. Type of object (Aircraft monoplane, toddy bowl)
3. For...
    a. Role: Shepperd
    b. Gender (Woman's)
4. Style/Technique (Heraldic panel, print - aquatint)
5. Place (Highlands, Paris)

S2 Named/Specific What (Action/Event/Condition)

Unique searches: 2

Total searches including frequencies/repetitions: 34

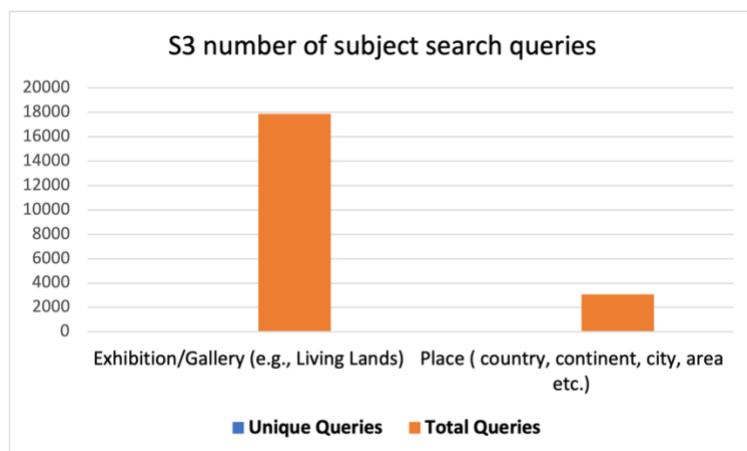| S2 Search Subject Queries	Unique Queries	Total Queries
Specific event (WW1, F1)	2	34
Total	**2**	**34**

S3 Named/Specific Where

Unique searches: 88

Total searches including frequencies/repetitions: 20 940

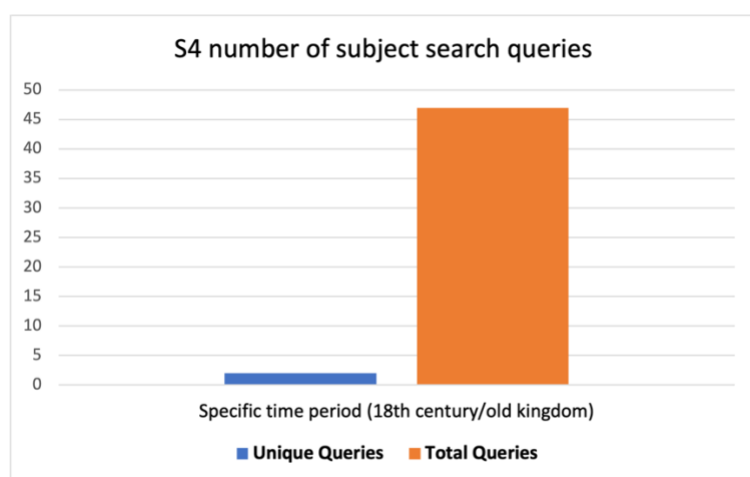| S3 Search Subject Queries	Unique Queries	Total Queries
1. Exhibition/Gallery (e.g., Living Lands)	34	17873
2. Place (mainly country but also continent, city, area and historical regions i.e., Mesopotamia)	54	3067
Total	**88**	**20940**

S4 Specific When/Time Period

Unique searches: 2

Total searches including frequencies/repetitions: 47

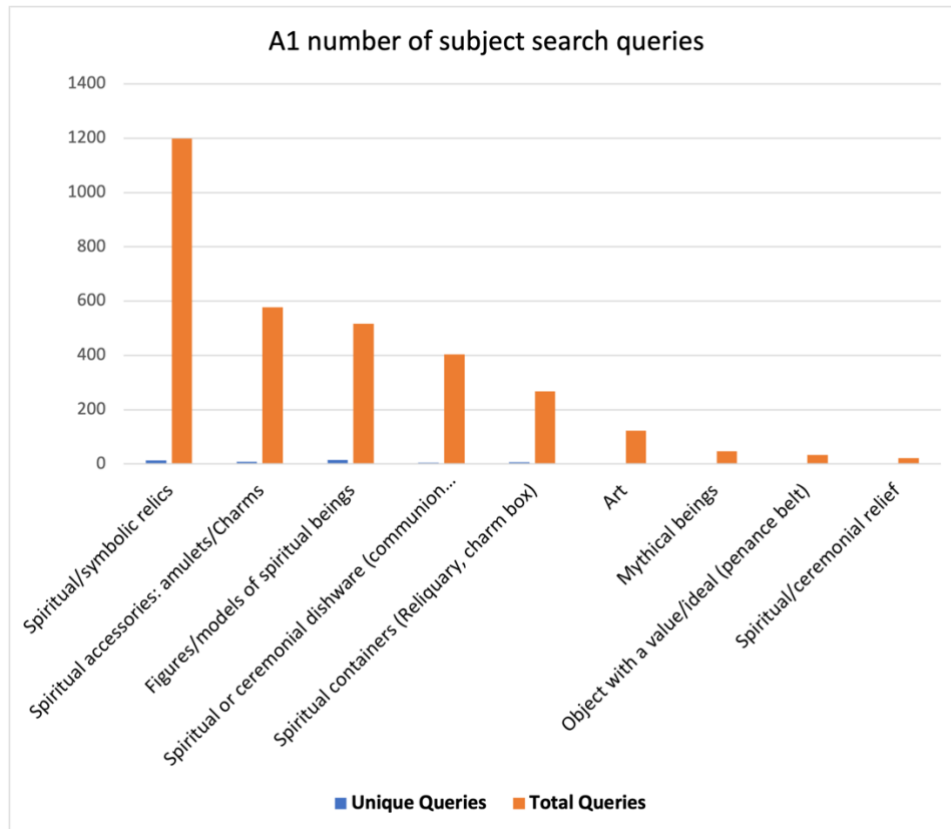| S4 Search Subject Queries	Unique Queries	Total Queries
Specific time period (18th century/old kingdom)	2	47
Total	**2**	**47**

A1 Abstract Who/Type of thing (Spiritual and mythical objects and people)

Unique searches: 56

Total searches including frequencies/repetitions: 3190

A1 Search Subject Queries	Unique Queries	Total Queries
1. Spiritual/symbolic relics (Magic knife, oracle bones, prayer wheel)	14	1199
2. Spiritual accessories: amulets/Charms (Amulet, Amulet stone, charm, Pons ikayop)	9	578
3. Figures/models/paintings of spiritual beings (Guardian figure, Water Nymph Figure, Surya Chariot horse figure, Thangka)	15	516
4. Spiritual or ceremonial dishware (communion cup/plate)	5	404
5. Spiritual containers, e.g., for relics/charms (Reliquary, charm box)	6	268
6. Art	2	122
7. Mythical beings	2	47

8. Object with a value/ideal (penance belt, propaganda vase)	2	34
9. Spiritual/ceremonial relief	1	22
Total	**56**	**3190**

Facets/Narrower for strings

- Representation of
- Type of object
- Container for...
- Material

A2 Abstract What (Action/Condition) & A3 Abstract Where (Spiritual places)

Unique searches: 5

Total searches including frequencies/repetitions: 265

1. Concept/Life event (Death)
2. Spiritual place (Church, shrine, Naos)

A2 & A3 Search Subject Queries	Unique Queries	Total Queries
Abstract/spiritual place	4	246

Abstract concept	1	19
Total	**5**	**265**