GETTY THESAURUS OF GEOGRAPHIC NAMES:
A Closer Look at the TGN

Patricia Harpring
Managing Editor, Getty Vocabulary Program
February 2012

TGN in Context
THE GETTY VOCABULARIES

- The Getty Vocabularies are authoritative “non-authoritarian” vocabularies used by the art and architecture community
- Data model allows linking to preferred names as well as to any other “non-preferred” name in the record
- CONA is the first Getty vocabulary to be linked to other vocabularies, although others have always used controlled lists based on each other (e.g., the place terms in ULAN are based on the TGN)
- Using the unique, persistent subject_id and term_id, Getty vocabularies may be used in linked data applications
- Vocabularies grow through contributions

http://www.getty.edu/research/tools/vocabularies/index.html

THE GETTY VOCABULARIES

- **Art & Architecture Thesaurus ® (AAT)**
  - 34,878 ‘records’; 245,531 terms
- **Getty Thesaurus of Geographic Names ® (TGN)**
  - 992,307 ‘records’; 1,711,115 names
- **Union List of Artist Names ® (ULAN)**
  - 202,718 ‘records’; 638,903 names
- **Cultural Objects Name Authority™ (CONA)**
  - *under development; now available for contributions*

- All four focus on the visual arts and architecture
- Grow through contributions from the user community (totals change daily)
- Compiled, maintained, distributed by the Getty Vocabulary Program, a unit of the Getty Research Institute (GRI) in Los Angeles
THE GETTY VOCABULARIES

- AAT terms = generic concepts
  - (e.g., watercolors, amphora)
- TGN names = administrative, physical places
  - (e.g., Los Angeles, Ottoman Empire, Bavarian Alps)
- ULAN = person, corporate body names
  - (e.g., Christopher Wren, Altobelli & Molins)
- CONA = titles/names of art and architecture
  - (e.g., Mona Lisa, Empire State Building)

- One main characteristic making the vocabularies useful as authorities is that each vocabulary record is identified by a unique, persistent numeric ID
- Meaning if a record changes or is merged, the original ID remains

The AAT, ULAN, and TGN were begun in the mid-1980s
- To meet the needs of the art museum, visual resources, archives, and art library communities
- The community needed authoritative vocabularies to aid in the indexing and retrieval of art information
  - To suggest a preferred term, while still allowing local institutional preferences regarding which term to use for the concept, person, or place
• Each vocabulary was conceived and constructed differently
• Since 1990s, all three have been united with the same core data structure and coordinated editorial rules
• Planning for CONA was begun in 2004
• Instruction for use of and contribution to the Getty vocabularies at conferences and online training materials
  www.getty.edu/research/tools/vocabularies/training.html

• Structure and content of the Getty vocabularies are based upon standards
  • ISO and NISO standards for thesaurus construction
  • Categories for the Description of Works of Art (CDWA) and Cataloging Cultural Objects (CCO) standards for art information
Thesauri

- Thesaurus: A semantic network of unique concepts
- Thesauri may be monolingual or multilingual
- Thesauri may have the following three relationships:
  - Equivalence Relationships
  - Hierarchical Relationships
  - Associative Relationships

Hierarchical

- Objects Facet
  - Furnishings and Equipment
    - Containers
      - <culinary containers>
      - <vessels for serving / consuming food>
- stirrup cups
- coaching glasses
- hunting glasses
- sturzbechers
- Sturzbecher
- stortebekers
- rhyta
- rhyton
- rhytons
- rhea
- rheon
- rheons
- ritón

Associative

- Objects Facet
  - Furnishings and Equipment
  - Containers
  - <culinary containers>
  - <vessels for serving / consuming food>

Equivalence

- The Getty vocabularies are used by various audiences
  - by catalogers or indexers who are describing works of art, archival materials, visual surrogates, or bibliographic materials
  - by researchers
  - by systems implementers creating search tools to enhance end-user access to online resources
HOW ARE VOCABULARIES USED?

- Used as sources of standard terminology for use in description, cataloging, and documentation
- Used as “assistants” in online search engines, creating a semantic “road map” that shows links and paths between concepts and terms
- Used as knowledge bases

Object type: Panathenaic amphora

Personal names:
- Qualifier: attributed to
  - Identity: Painter of the Wedding Procession
    (Greek vase painter, active ca. 360s BCE) Role: painter
  - Identity: Nikodemos
    (Attic potter, active ca. 362 BCE) Role: potter

Materials and techniques:
- Material Name: terracotta
- Technique Names: turning, sintering, vase painting

Geographic names:
- Creation Place/Original Location: Athens (Periféreia), Protevounai, Greece

Corporate names:
- Repository Name Geographic Location: Getty Villa Malibu, J. Paul Getty Museum (Los Angeles, California, United States)

Iconographic subjects:
- Extent: side A
  - Indexing Terms: Athena Promachos, human female
- Extent: side B
  - Indexing Terms: Athena/Parthenos, ceremonial object
HOW ARE VOCABULARIES USED?

• There are over 300 licenses for the Getty vocabularies, which may each represent a large number of users of a vendor’s system
  • Data is accessible in annual releases of XML and relational tables and through Web Services, where the data is refreshed every two weeks
  • In addition, the online Web presentation of the vocabularies typically supports around 180,000 sessions by users each month
    • The Getty vocabularies are the most frequently used online resource at the Getty Research Institute, and among the most heavily accessed at Getty as a whole

HOW ARE VOCABULARIES USED?

• Getty vocabulary terms and associated information are valued as authoritative because they are derived from published sources and represent current research and usage in the art history community.
  • The vocabulary program publishes extensive editorial manuals online. We also make frequent presentations and conduct training workshops at the Getty and at conferences throughout the year.
CONTRIBUTIONS

- The Getty vocabularies are compiled in large part from contributions from the user community, including various Getty projects and qualified outside institutions.
- Institutions may make contributions in bulk in our prescribed XML format or via an online form.
- Bulk contributions can range from a few thousand to several million records.

CONTRIBUTIONS

- Contributions to our vocabularies are vetted, managed, edited, augmented, merged with existing data, placed into hierarchies and other relationships, and published by the Getty Vocabulary Program.
- Licensed files are released annually; the data on the Web site is refreshed every two weeks.
• Getty vocabularies are increasingly multilingual
  
  • The Getty vocabularies have terms/names/titles and descriptive notes in multiple languages
  
  • TGN and ULAN have many thousands of non-English names, although the languages are not always flagged by contributors and thus cannot be counted
  
  • AAT: Of the approximately 250,000 total terms in the AAT, about 116,000 are not American English: full translations in Spanish and Dutch are incorporated; 6,000 terms have French and Italian equivalents; 3,000 are British English; 1,600 are German (statistics from late 2011)
simplified Entity Relationship Diagram for Getty Vocabularies

SUBJECT
MAIN TABLE
basic record information, unique ID, parent_key, record type, descriptive/scope note, flags

NAMES / TERMS
multiple names, one is flagged preferred; dates for names

ASSOCIATIVE RELATIONSHIPS
links between subjects, dates

• Equivalence relationships between terms/names that are linked to the same concept ID

• Hierarchical relationships between different concept IDs; each record is linked to its immediate parent

• Associative relationships between different concept IDs; are reciprocal

PLACE TYPES/ROLES
multiple place types or roles, one is flagged preferred, dates

Biography
display bio, birth/death dates, places

Revision History
tools

Sources

Contributors

Coordinates

Events
date, places

Languages

Nationality

TGN, ULAN, and CONA have additional tables
simplified Entity Relationship Diagram for Getty Vocabularies

**SUBJECT**
- MAIN TABLE
  - basic record information, unique ID, parent_key, record type, descriptive/scope note, flags

**NAMES / TERMS**
- multiple names, one is flagged preferred; dates for names

**ASSOCIATIVE RELATIONSHIPS**
- links between subjects, dates

**Controlled values**
- identified by IDs, other shorter controlled lists in the DB simply list terminology (e.g., "preferred," "non-preferred")
- CONA is first to actually link to the other vocabularies for some fields, rather than maintaining refreshable lists

**Linked to Getty Vocabularies**
- CONA is the first Getty vocabulary to be actually linked to the others (others have controlled lists, albeit based on another vocabulary)

**Authorities**
- ULAN
- TGN
- AAT
- Iconography Authority

**CONA Work Records**

**Source Records**
CONA Record

CONA ID: 700000153

Class: Sculpture  Work Type: hacha
Title: Hacha (Ceremonial Ax)
Creator Display: unknown Veracruz
Creation Date: Late Classic Veracruz (600-900 CE)
Current Location: Dumbarton Oaks (Washington DC, USA)  ID: 838 VCS

Discovery Location: Veracruz (Mexico)

Subject: • religion/mythology  • human figure  • face  • game  • sacrifice  • ceremonial object  • Inca ballgame

Measurements: 35 x 22.2 cm  (height)  (13 3/4  x 8 3/4  inches)
Materials and Techniques: marble

Descriptive Note: The hacha had ritualistic significance and likely served as an actual ballcourt marker.

Getty Thesaurus of Geographic Names
Details of TGN
ELEMENTS OF A TGN RECORD

names
Munich
München
Monaco
Munich

parent place
Germany
Bavaria
Oberbayern

place types
inhabited place
state capital

coordinates
48°08'48"N, 011°35'01"E

dates
founded near an older settlement in 1157

note
Capital of Bavaria and the third-largest city in Germany; is situated on both sides of the Isar River, north of the Alps. Henry the Lion, duke of Bavaria, established it in 1157 as a mint and market for Benedictine monks from Tegernsee...

sources

---

TGN is a thesaurus

- Thesaurus: A semantic network of unique concepts
- Thesauri may be monolingual or multilingual
- Thesauri may have the following three relationships:
  - Equivalence Relationships
  - Hierarchical Relationships
  - Associative Relationships

La Piana (deserted settlement)  hierarchical
  ....  World (facet)
  .......  Italy (nation)
  ..........  Tuscany (region)
  ............  Siena province

Siena
Σιένα
Siyenah
锡耶纳
Сиена
ה夌
Siensce
Sienna
Saena Julia
Sanna
Sena
Senae

Pisa (inhabited place)  equivalence

associative
adjacent to
ally of
SCOPE OF TGN

- TGN is a thesaurus, compliant with ISO and NISO standards for thesaurus construction
- It contains hierarchical, equivalence, and associative relationships
- Is TGN a GIS (Geographic Information System)? No. A GIS is designed to capture, store, manipulate, analyze, manage, and present geographically referenced data; coordinates are key
- TGN focuses instead on names and relationships appropriate to a thesaurus
- While many records in TGN include coordinates, these coordinates are approximate and are intended for reference only; TGN is not linked to or capable of constructing maps
- However, TGN could be used with GIS to link to maps and to improve retrieval of GIS based on relationships, including synonyms (variant names)

SCOPE OF TGN

New places in TGN must be within scope

- TGN places include political entities and physical features
- Scope is global, some extraterrestrial
- Includes all current continents and nations
- Historical places, including nations and empires
- Prehistory to the present
- Real places, not mythical
- May include formerly inhabited places, historical places with unknown exact locations “lost settlement”
- Focus on places important to art and material culture
### Required Fields for TGN

- preferred name
- variant names, as stated in Rules
- source(s) for the names
- place type(s) (e.g., inhabited place)
- hierarchical position

### List of Major Fields for TGN in VCS editorial system

#### 3.1 HIERARCHICAL RELATIONSHIPS
- Parents (required)
- Sort Order (required)
- Historical Flag (required)
- Dates for relationship to parents
- Hierarchy Relationship Type (required)

#### 3.2 IDENTIFYING NUMBERS, STATUS FLAGS, AND DESCRIPTIONS
- Subject ID (required)
- Parent Key (required)
- Merged Status (required)
- Published Status (required)
- Record Type (required)
- Candidate Status (required)
- Label (required)
- Contributors for Subject Record (required)
- Sources for the Subject Record (required)

#### 3.3 NAMES
- Term ID (required)
- Name (required)
- Preferred Flag (required)
- Qualifier
- Sequence Number (required)
- Historical Flag (required)
- Term Type (required)
- Part of Speech (required)
- Vernacular Flag (required)
- Language for Names (required)
- Preferred Flag for Language (required)
- Language Status (required)
- Contributor for Name (required)
- Preferred Flag for Contributor (required)
- Sources for Names (required)
- Page Number for Term Source (required)
- Preferred Flag for Source (required)
- Dates for Names
- Display Name Flag (required)
- AACR Flag (LC heading)
- Other Flags
- Assigned To note

*Data Dictionary for release formats are available from links on this page
*http://www.getty.edu/research/tools/vocabularies/obtain/download.html
List of Major Fields for TGN in VCS editorial system

3.4 DESCRIPTIVE NOTE
- Descriptive Note
- Sources for the Descriptive Note
- Contributors for the Descriptive Note
- Language of Descriptive Note

3.5 ASSOCIATIVE RELATIONSHIPS
- Related Places
- Relationship Type
- Historical Flag
- Dates for Associative Relationship

3.6 PLACE TYPE
- Place Type (required)
- Preferred Flag (required-default)
- Sequence Number (required-default)
- Historical Flag (required-default)
- Dates for Place Type

3.7 COORDINATES
- Coordinates
  - Latitude: Degree; Minute; Second; Direction; Decimal Degrees
  - Longitude: Degree; Minute; Second; Direction; Decimal Degrees
- Bounding Coordinates
  - Least Latitude: Degree; Minute; Second; Direction; Decimal Degrees
  - Most Latitude: Degree; Minute; Second; Direction; Decimal Degrees
  - Least Longitude: Degree; Minute; Second; Direction; Decimal Degrees
  - Most Longitude: Degree; Minute; Second; Direction; Decimal Degrees
- Elevation: Feet; Meters

3.8 ADMINISTRATIVE FLAGS, NOTES, AND REVISION HISTORY
- Comment Flag
- Problem Flag
- Assigned To
- Special Project
- Facet
- Legacy ID
- Class Notation
- Image
- Index Note
- Not Found Note
- Status Note
- Editor Note
- Revision History (required-default)
### WHAT IS A NAME IN TGN

- Names, appellations, and designations used to identify a place
  - Current, historical, official, local usage
  - May include a “core” name and descriptive designation (e.g., Mount Etna or Mississippi River)
- MUST be equivalents, referring to the same place
- If an archaeological site is near but not on the site of a modern town, these are two separate records, not two names in the same record

### EXAMPLES OF PLACE NAMES

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firenze</td>
<td>vernacular names</td>
</tr>
<tr>
<td>Florence</td>
<td>variant names in other languages (English)</td>
</tr>
<tr>
<td>Fiorenza</td>
<td>historical names</td>
</tr>
<tr>
<td>Tel Hazor</td>
<td>transliterations</td>
</tr>
<tr>
<td>Tel-Hazor</td>
<td>variations in spelling, diacritics, punctuation, or capitalization</td>
</tr>
<tr>
<td>Big Apple</td>
<td>nicknames</td>
</tr>
<tr>
<td>City of Brotherly Love</td>
<td>translations</td>
</tr>
<tr>
<td>Michigan, Lake</td>
<td>inversions (physical features)</td>
</tr>
<tr>
<td>Florentine</td>
<td>adjectival forms</td>
</tr>
</tbody>
</table>
PREFERRED NAMES

- One name flagged “preferred”
- “Preferred” name is the name most commonly used
- Generally vernacular; but English for continents, oceans, historical entities, etc.

- Chosen from authoritative scholarly sources and general reference works in American English
- Include English variant if there is one, flag it
- May have been transliterated into Roman alphabet by your source where necessary
- Please include variant names as well

Equivalence Relationships

Mississippi River
Mississippi
Mississippi, fleuve
Fiume del Missisipi
Fleuve Mississippim
Chicagua
Chucagua
La Grande Riviere
Malabanchia
Malabouchia
Mascicciipi
Meact-Chassipi

- one name is required
- many records have multiple names
### Equivalence Relationships

**Names:**
- Athinaí (preferred, C,V,Greek (transliterated), U)
- Athénai (C,O)
- Athinaí (C,O,Greek (transliterated), U)
- Άθηναι (C,V,Greek, U)
- Athen (C,O,English-P, U,N)
- Athenas (C,O,Spanish-P, U,N)
- Αθήναι (C,O,French-P, U,N)
- Athen (C,O,German-P, U,N)
- Athens (C,O,Prefered, U)
- Άθηναι (C,V,Ancient Greek, U)
- Aithina (NA,V,Ru)
- Athina (NA,V,Gr)
- Affini (NA,V,Fr)
- Aten (NA,V,Per)
- Aten (NA,V,Sw)
- Atény (NA,V,Po)
- Athenés (NA,V)
- Athni (NA,V)
- Athina (NA,V)
- Athenas (NA,V)
- Αθηνα (NA,V)
- Αθηναί (NA,V)
- Atenas (C,O,Spanish-P, U,N)
- Athenes (C,O, French-P, U,N)
- Athenai (H,V,Prefered, U)
- Athenai (H,V,Ancient Greek (transliterated), U)
- Athine (H,V,Greek (transliterated), U)

- **one name is required**
- **many records have multiple names**
- **(online display)**

---

### NAMES IN TGN

- Firenze (preferred, vernacular, Italian-preferred)
- Florence (English-preferred)
- Florencia (Spanish-preferred)
- Florenz (German-preferred)
- Fiorenza (historical)
- Medieval
- Florentia (historical, Latin) name of Roman colony on N bank of Arno
- Florentine (adjectival, English)

- **All names MUST refer to the same place (e.g., lost sites)**
- **Record names in mixed case (not all caps)**
- **Generally only 2 to 4, not more than 15 names (obscure archival names not mandatory or desirable, but published names are desired)**

---

Patricia Harpring; copyright J. Paul Getty Trust; February 2012
### NAMES IN TGN

- **Firenze (preferred, vernacular, Italian-preferred)**
- **Florence (English-preferred)**
- **Florenzia (Spanish-preferred)**
- **Florenz (German-preferred)**
- **Fiorenza (historical, Medieval)**
- **Florentia (historical, Latin) name of Roman colony on N bank of Arno**
- **Florentine (adjectival, English)**

- Prefer the vernacular form commonly used in standard authoritative sources in English language
- Include the preferred English form as well when it differs from vernacular
  - generally only for famous places and large places (nations, mountain ranges)

### Multiple Vernacular Names

- **Strasbourg (preferred, vernacular, French-preferred)**
- **Strassburg (vernacular, German-preferred)**
- **Straßburg (vernacular, German-preferred)**
- **Strossburi (vernacular, Alsatian-preferred)**
- **Estrasburgo (other, Spanish-preferred)**
- **Strateburgum (historical, Medieval Latin)**
- **Argentoratum (historical, Ancient Latin)**

- multiple vernacular languages
- important to record all
- prefer the one used in standard sources that also prefer the vernacular (e.g., atlas, gazetteer, NGA/NIMA)
Transliterations, diacritics

Tōkyō (preferred, C,V) ............ after replacing
Kyōto as Imperial capital, 1868
Tokyo (C,V, English-H, U, N)
東京 (C,V, Japanese-P, V, N)
Tokio (H,O, English, U, N) ............ obsolete spelling
Edo (H,V) ............ during Tokugawa Shogunate
beginning 1603
Yedo (H,V)
Yedo (H,V)

Processing diacritics

T'ainan (preferred, C,V,N)
Tai-nan (C,V,N)
T'ai-nan (C,V,N)
T'ai-nan-shih (C,V,N)
Tainanshi (C,V,N)
Tainanshih (C,V,N)
臺南市 (C,V,N)
Dainan (H,O,N)
T'ai-wan (H,V,N)
T'ai-yüan (H,V,N)
Ta-yüan (H,V,N)

- Transliterated into the Roman alphabet
- Generally using a transliterated source, but if possible, prefer ISO
- ISO standards provide the preferred transliteration protocol (Pinyin, but include Wade-Giles for access too)
Display Name, Indexing Name

Naples (Napoli province, Campania, Italy, Europe)

Napoli (vernacular preferred)
Naples (English-preferred)
Napoli province (display name)

• Flag "display name" = special name flagged for horizontal displays or "headings"
• where simple name is confusing due to redundancy, etc.

Names and Language Flags

al-Urdunn (Vernacular, preferred)
Al Urdunn (Vernacular)
Urdunn (Vernacular)
al-Mamlakah al-Urdunniah al-Hāshimiyah (Vernacular, official)
Jordan (preferred English)
Hashemite Kingdom of Jordan (English)
Jordanie (French)
Royaume hachémite de Jordanie (French)
Jordaan (Dutch)
Jordania (Spanish)
Giordania (Italian)
JOR (ISO 3L)
JO (ISO2L)
ISO400 (ISO3N)
J000 (FPS)
Transjordan (Historical) Display
April 1949 Earliest: 1921 Latest: 1949

• vernacular and other languages
• variant transliterations
• preferred English, ISO, and other flags
• preferred vs. official name, not necessarily the longest form
Names and date

Siena (vernacular, preferred) used since 13th century (start: 1200, end: 9999)
Sienna (English) obsolete spelling (start: 1500, end: 1900)
Senac (historical) medieval (start: 800, end: 1500)
Sanna (historical)
Saena Julia (historical) Roman (start: -100, end: 300)
Sena Julia (historical)
Sena (historical) Etruscan (start: -800, end: -100)

- names may be current or historical
- names may have display dates
- years delimiting span are included for retrieval, not displayed to end user

Hierarchical Relationship in TGN - Whole/Part

Top of the TGN hierarchy (hierarchy root)
.... World (facet)
....... North and Central America (continent)
............ Costa Rica (nation)
................. Alajuela (province)
................. Cartago (province)
................. Guanacaste (province)
................. Heredia (province)
................. Limon (province)
................. Puntarenas (province)
................. San Jose (province)

- provinces are part of the nation
- displayed with indentation
- indicated by linking each place to its “parent”
• Choosing the parent
• To add a new record, TGN requires that you name the correct administrative division within the nation where the place should be positioned
• Consult the Rules and precedent for other places in that nation
• If you cannot determine the correct administrative subdivision, use the narrowest level known (e.g., state or nation, if you do not know the county)

Levels below inhabited place

- hierarchy generally descends to level of inhabited place
- neighborhoods are included for largest cities

World
- North and Central America (continent)
- United States (nation)
- California (state)
  - San Francisco (inhabited place)
    - Chinatown (neighborhood)
    - Ingleside (neighborhood)
    - Mission (neighborhood)
    - Nob Hill (neighborhood)
    - North Beach (neighborhood)
    - Washerwomans Bay (former physical feature)
Historical and former places

- Some places may no longer exist, e.g., submerged island of Alexandria, Egypt

Polyhierarchy

- Multiple hierarchical relationships may include historical parents
- The date of the relationship may also be included
**Dates for hierarchy**

Hierarchical Position:
- World (facet)
  - Europe (continent)
    - Italy (nation)
    - Umbria (region)
    - Terni (province)

- The date of the relationship may also be included.

E.g., Orvieto is linked to the historical entity, Etruria, with dates.

**Changing relationships through history**

- Urban diminishment
- Occasionally large urban areas are replaced with several smaller towns
- E.g., villages on site of deserted settlement, Thebes

Top of the TGN hierarchy (hierarchy root)
- World (facet)
  - Europe (continent)
    - Italy (nation)
    - Umbria (region)
    - Terni (province)
  - Etruria (former group of nations/status/cities) (H)
    - Orvieto (inhabited place) (H) as Velzna, one of the major cities of the Etruscan foes destroyed in 264 BCE.
Urban expansion

Former independent towns incorporated into modern city, Vienna

World
Europe (continent)
Austria (nation)
Vienna (state)
Vienna (inhabited place)
Almannsdorf (suburb)
Aspern (suburb)
Atzgersdorf (suburb)
Breitenlee (suburb)
Dornbach (suburb)
[...]
Stammersdorf (suburb)
Strebersdorf (suburb)

Physical feature crosses national boundaries

- physical features that cross boundaries are placed under the level that entirely contains them
- Reasons: focus of TGN is on inhabited places; complex linking of physical features would clutter the thesaurus

World
Europe (continent)
Adriatic Sea (sea)
Alemannia (region, general)
Alps (mountain system)
• Alsace-Lorraine (region, general)
• Andorra (nation)
- no set number of levels in administrative structure
- Different nations have different numbers of levels
- Also, to facilitate island nations, TGN has the policy of incorporating the physical features, islands, into the hierarchy, even if they are not coextensive with administrative division

E.g., hierarchy with physical feature as intervening level, Pate Island, Kenya

Africa (continent)
Kenya (nation)
Coast (province)
  Bura (inhabited place)
  Galana (river)
  Gazi (inhabited place)
  Kinango (inhabited place)
  Lamu (inhabited place)
  Lugards Falls (waterfalls)
  Mkunumbi (inhabited place)

Pate Island (island)
  Pate (inhabited place)
  Rasini (inhabited place)
  Rabai (inhabited place)
  Shimoni (inhabited place)

General region, Middle East

Note - Extensive area, including nations of SW Asia & NE Africa; term formerly used to also include Afghanistan, Pakistan, India & Burma; term "Near East" is used to refer to the same general area as "Middle East," but may also be used to mean a larger area.

Names:
  Middle East (C,O)
  Le Moyen-Orient (C,O)
  Oriente Médio (C,O)
  El Medio Oriente (C,O)
  Vorderasien (C,O)
  Near East (C,O)
  Proche-Orient (C,O)

World
  Africa............(continent)
  Antarctica.....(continent)
  Asia..............(continent)
  Central Asia...(region (general))
  Eurasia.........(region (general))
  Europe..........(continent)
  Middle East (region, general)
  North and Central America
  Oceania.........(continent)
  South America.(continent)

"General region" = no consistently defined boundaries
 currently has no hierarchical depth
Place types for Indianapolis, Indiana

inhabited place
city
state capital
manufacturing center
transportation center
sporting center
financial center
agricultural center

- minimum record has one place type
- fuller records have more
- one is flagged as “preferred”

- For cities, towns, villages, preferred = inhabited place

Place types for Indianapolis, Indiana

inhabited place ... founded in 1821
Start: 1821   End: 9999

city
state capital ... since 1825 Start: 1825 End: 9999
manufacturing center
transportation center
sporting center ... especially noted for Indianapolis 500 automobile race, since 1911 Start: 1911 End: 9999
financial center
agricultural center

- May have dates with place type
- Dates of habitation are most important
Place types for Machupicchu, Peru

- Historical or current
- "deserted settlement" is preferred current, "inhabited place" is historical

Deserted settlement (preferred, current) ... building started ca. 1440; was inhabited until the Spanish conquest of Peru in 1532
  Start: 1430 End: 1550
Archaeological site (current) ... rediscovered in 1911
  Start: 1911 End: 9999
Ruins (current)
Inhabited place (historical)
Inca center (historical) ... building started ca. 1440; was inhabited until the Spanish conquest of Peru in 1532
  Start: 1440 End: 1550

Choose appropriate Place Type from the controlled list.
ASSOCIATIVE RELATIONSHIP IN TGN

**Hazor**

**Tel Hazor**

- There may be associative relationships between places
  - deserted settlement may be related to a modern town
  - date for relationship is possible

for a deserted settlement in Israel

**DESCRIPTIVE NOTE:**
Located over Huleh Valley, near the modern town Hazor; for centuries it was one of most important cities of Canaan, in control of rich agricultural area and vital trade and military route; according to the Bible, site of victory of Joshua and other battles of Israelites.

### Associative Relationship examples

- **[for the Ancient Mesopotamian kingdom, Assyria, which is distinct from the Roman Province of the same name]**
  - Relationship Type: distinguished from
  - Related Place: Assyria (Roman Empire) (province)

- **[in the record for the lost settlement of Sharuhen]**
  - Relationship Type: possibly identified as
  - Related Place: Tel el-Far'ah (As Suwayd$01a', Syria) (deserted settlement)

- **[for Kings county, New York, USA]**
  - Relationship Type: coextensive with
  - Related Place: Brooklyn (New York, New York, USA) (borough)

- **[for ancient Persia]**
  - Relationship Type: predecessor of
  - Related Place: Iran (nation)
**Associative Relationship examples**

- **Dates for Relationships**

  [for Lincoln, England]
  Relationship Type: capital of
  Related Place: Flavia Caesariensis (Britannia Inferior, Roman Republic and Empire) (province)
  Display Date: from the early 4th century CE
  Start Date: 296  End Date: 900

  [for Florence, Italy]
  Relationship Type: ally of
  Related Place: Orvieto (Terni province, Umbria, Italy) (inhabited place)
  Display Date: Guelf allies during the 13th and 14th centuries
  Start Date: 1250  End Date: 1400

---

**COORDINATES**

- Derived from published sources: atlases, NGA/NIMA, USGS
- For “finding purposes,” not GIS
- Expressed in degrees, minutes, and seconds
- Translated into decimal degrees
- Represent a point at the center of a place, or the source of linear features (e.g., rivers)
- Elevation
- Bounding coordinates

### COORDINATES

- **Latitude Degree**
- **Minute**
- **Second**
- **Direction**
  - N
  - S

- **Longitude Degree**
- **Minute**
- **Second**
- **Direction**
  - E
  - W

- **Elevation Feet**
- **Elevation Meters**

**Click to add bounding coordinates**
Recording a single point
- Usually adequate for smaller places
- Point = source of linear features
- Marginally helpful for center of areas (may record point AND bounding box)

Coordinates for nation of India
Latitude: 20 00 N, Longitude: 077 00 E

Descriptive Note
For Gaul (historical region)
Note: Refers to the region inhabited by the ancient Gauls, comprising modern-day France and parts of Belgium, western Germany, and northern Italy. It was a powerful ancient country. Inhabited from ca. 600 BCE by Celtic Galli. French Gaul (the area of Gaul that is modern France) was later divided by Rome into four provinces: Narbonensis, Aquitania to the west and south of the Loire, Lugdunensis in central France between the Loire and the Seine, and Belgica in the north and east.

Topics may include the following:
- disputed issues, sovereignty, or ambiguity regarding the names or critical facts about the place
- possible confusion of the place due to its name being a homograph for a nearby place
- physical description
- description of its location (do NOT repeat the hierarchy information, but you may mention physical features, etc.)
- the first habitation of the place
- its political history, in chronological order
- its importance relative to other places or to the history of art and architecture
- for modern nations, include the languages spoken
- population may be included for inhabited places, provided you cite the date of your source

map from http://www.uni-karlsruhe.de/~uk4u/indiamap.html
Sources

- Required to list sources.
- For preferred name, prefer the most authoritative, up-to-date sources available.
- Sources of information in the TGN record may include the following in this order of preference:

  **Standard general reference sources**
  - atlases, loose maps, gazetteers
  - geographic dictionaries, encyclopedias, guidebooks
  - government Web sites, including NGA (NIMA) and USGS

  **Other official sources**
  - newsletters from ISO and United Nations
  - communication with embassies
  - Library of Congress subject headings

  **Other material on topics of geography or current events**
  - books, journal articles, and newspaper articles
  - archives and other original sources

  **Other sources**
  - inscriptions on art objects, coins, or other artifacts
  - catalog records of repositories of art objects
  - books on history of art and architecture

Indexing with TGN
Use historical names?

Unique TGN IDs exist for both place and each name

Class: sculpture
Work Type: sculpture, votive figures
Title: Ploughman of Arretium
Measurements: 7 1/4 in. length
Material: bronze
Creator: unknown Etruscan
(link) unknown Etruscan Role: sculptor
Creation Date: 6th century BCE

Creation Location: Arretium

Style: Etruscan
Subject: ploughman, oxen, labor, afterlife, ceremonies

Description: This figure was found in one of the many sanctuaries from Etruscan Arretium, and was probably a votive offering. The subject reflects a natural preoccupation of the inhabitants of Arretium, which was an outpost of Chiusi in a fertile farming region.

Current Location: Archaeological Museum (Florence, Italy)

Location Authority
Arezzo (preferred, Italian-Pref)
Arretium (historical, Ancient Latin)

If the historical entity is not the same place, concept, person, etc. as the modern counterpart
- Then the two names/terms should be in separate authority records
- Thus you should link to the correct historical entity
- E.g., historical Flanders did not have the same boundaries as modern Belgium; they are NOT equivalents in a thesaurus
Processing and Using TGN

How to disambiguate homographs in mapping to CONA

When loading contributions into CONA, how to automatically choose the TGN homograph intended, if the broader context is not given in the contributed record?

1. Editor interaction
2. In some cases, such as place of publication, automatically assume the famous “Boston” Massachusetts is intended. It is flagged as such in TGN to allow auto-matching.

- Boston (inhabited place) (World, Asia, Philippines, Davey Oriental) [1084507]
- Boston (inhabited place) (World, Europe, United Kingdom, England, Lincolnshire) [7010726]
- Boston (inhabited place) (World, North and Central America, United States, Alabama, Covington county) [2158205]
- Boston (inhabited place) (World, North and Central America, United States, Arkansas, Madison county) [2007790]
- Boston (inhabited place) (World, North and Central America, United States, Connecticut, Fairfield county) [201705]
- Boston (inhabited place) (World, North and Central America, United States, Georgia, Thomas county) [2022223]
- Boston (inhabited place) (World, North and Central America, United States, Georgia, Thomas county) [2022223]
How to disambiguate homographs
merging duplicates in TGN loads

During loads of contributed data in TGN, records are merged automatically if names, coordinates, broader context, and place types match or closely correspond.

Are these two records for the same place? If so, they will be “merged,” automatically if the data supports it, or by hand if data requires research.

For loading contributions:
Reports indicate to editors which possible matches require human research.
E.g., all compared fields match, except one: place type does not match.
How to disambiguate homographs in displays for end users

Alexandria (Asia, Pakistan, Baluchistan)......(deserted settlement)

Alexandria (Baluchistan, Pakistan) (deserted settlement)

- Parent string in reverse order allows indexing results lists by continent, nation
- In natural order is easier for user to read

Accommodate requirements of horizontal vs. hierarchical displays. E.g., “display name” for parent string

- special name flagged for horizontal displays
- where simple name is confusing
- in a heading

Names:

Coast (preferred, C,V,N)
Coast Province (display, C,V,N)

Galana (Coast Province, Kenya, Africa) river
Hierarchical display for thesauri

- Preferred display of hierarchies is with indentation for a general audience, not thesaurus specialists
- Use graphics to indicate focus and where levels may be expanded

English display

- Top of the TGN hierarchy (hierarchy root)
- ... World (facet)
- ....... Europe (continent)
- ........... Greece (nation)
- ......... Aegean Islands (region)
- .......... Sporades (department)
- .......... Rhodes (island)
- ................ [ View physical features ]
- .................................................. Aithou (inhabited place)
- .................................................. Apollonia (inhabited place)
- .................................................. Aithou (inhabited place)
- .................................................. Asklepios (inhabited place)
- .................................................. Kameiros (deserted settlement)

Vernacular display

- Top of the TGN hierarchy (hierarchy root)
- ... World (facet)
- ....... Europe (continent)
- ........... Ellas (nation)
- ........... Nisos Ayvalik (region)
- ............ Sporades (department)
- .......... Rodes, Nisos (island)
- ................ [ View physical features ]
- .................................................. Aithou (inhabited place)
- .................................................. Apollonia (inhabited place)
- .................................................. Aithou (inhabited place)
- .................................................. Asklepios (inhabited place)
- .................................................. Kameiros (deserted settlement)

Allow display of the hierarchy in multiple languages; TGN currently supports English and vernacular

English view may be constructed by using flagged English name
TGN in Retrieval

- How to search by name / term?
  - full name string, exact
  - full name string, normalized
  - truncation
  - key words
  - Boolean operators
Tools such as Soundex and other phonetic algorithms do not work well with geographic or personal names and most AAT terms, because all languages (not only one language) may be represented in the vocabularies.

Therefore, specialized normalization and searching algorithms are advised for Getty vocabularies.

FULL NAME STRING
EXAMPLES
Florenz
Rancho San Diego
River Thames

ISSUES
• Can yield accurate retrieval, important for large data sets & for finding errors
• Queries fail due to no exact match (e.g., Thames or Thames, River)
• Query on normalized name increases hits (e.g., Rancho Sandiego); remove case, spaces, punctuation, diacritics

KEY WORDS
EXAMPLES
Thames
Diego
Sena

ISSUES
• Allows retrieval when exact format of multiple word names is unknown (e.g., Thames, or River Thames, or Thames, River)
• Causes problems for compound words if full name query is not also an option (e.g., Rancho Sandiego vs. San Diego, Rancho)
• Most effective with Boolean operators an option, otherwise results list may be too long
How results can differ

**TRUNCATION**

**EXAMPLES**
- Edinburg*
- Pyeitawinzu Myanmar* Nain*
- *angqiao

**ISSUES**
- Wild card allows retrieval when exact spelling is unknown
- Left-hand truncation is very helpful, but impractical with large data sets
- Wild card may be applied to full name string and to key words

**BOOLEAN OPERATORS**

**EXAMPLE: full name**
- Bay of Biscay OR Biscay, Bay of

**EXAMPLES: key words**
- (Castillo OR Rancho) AND Diego
- Monte AND Oliv*

**ISSUES**
- Allows best retrieval, most control by user
- May be applied to full name string and to key words
- Naïve user may not understand logic of formula (AND & OR, parentheses)
- Place name may be homograph of operator, e.g. “Or”

How results can differ

**STOP LIST**

**EXAMPLES**
- West
- Lake
- al

**ISSUES**
- Not the same as general database (e.g., not articles & prepositions)
- Should be allowed with Boolean operators, or results will be too limited
- Need way to allow queries on names composed only of stop words (e.g., town called “Lake”)

**ABBREVIATIONS**

**EXAMPLES**
- St. Louis
- W Lafayette
- Mt Everest
- Moskovskaya Ob

**ISSUES**
- Map to full word allows best retrieval, but difficult for vernaculars
- Often affects place types in names, directional indicators
Key Words and Boolean

- How we achieve good retrieval: special processing for searching
- Names in natural or inverted order (pivot using the comma)
- Retrieval by keywords or normalized strings
- Automatic “anding” or “oring” is not successful because it brings back too many results

1. **Ghunayjah, Jabal al** ........... (mountain) (World, Africa, Egypt, Desert, Shamāl Sināʾ) [7024974]

   Possible queries from users:
   - Ghunayjah, Jabal al-
   - Ghunayjah
   - Jabal AND Ghunayjah
   - Jabal Alghunayjah

   “Keyword” table created for matching:
   - ghunayjah
   - jabal
   - al
   - ghunayjahjabalal
   - jabalalghunayjah

Normalizing diacritics

- Accommodating diacritics
  - in query
  - in data (Unicode, other codes)
  - if display of non-Latin 1 not possible

   **Oriente** ........... (region)
   (World, South America, Ecuador) [7024091]
   Región Amazónica
   Ecuador Amazónico

   **Grande Região Norte** ........... (region)
   (World, South America, Brazil) [7023444]
   Amazônia Brasileira
   Variant Amazônica, Região

   **Amazon Basin** ...........(basin)
   (World, South America) [7024123]
   Hoya Amazônica
   Bacia Amazônica

   “Keyword” table is normalized, omitting diacritics:
   - amazonica
   - amazonica

   TGN data base:
   - Amaz$00onica
   - Amaz$03onica
Other criteria for search:
- Browsing hierarchies
- TGN: Place type, broader context (nation or continent), [could add coordinates, but we do not currently include them here]

Issues are involved with incorporating vocabularies into the user searching experience (as of February 2012)
- We are still working out the UI for one issue, for TGN and the other vocabularies on our own online materials
- For art history and related disciplines, the materials being searched do not necessarily include the broader context for a place name (at least not in a way that can be extracted by a computer; it may be apparent to a human reading it), because they are not linked to a geographic system
- Case study: if a user searches for "Los Angeles" in target materials, TGN can successfully offer 70 TGN records containing that name along with all their variants, including the famous California city as well as a city in Chiapas, Mexico, and others
- However, if the materials being searched do not indicate the broader context "Mexico," even if the user narrows the search to "Los Angeles, Mexico" he/she will still be overwhelmed by results including "Los Angeles" (California), or (depending on how developers calibrate the results) retrieve zero or incomplete results if only materials including a reference or context of "Mexico" are retrieved