

# From Assets to Stories via the Google Cultural Institute Platform

W. Brent Seales  
*Professor, Computer Science*  
*University of Kentucky*  
*Lexington, Kentucky, USA*  
*seales@uky.edu*

Steve Crossan  
*Founder, Cultural Institute*  
*Google*  
*Paris, France*  
*scrossan@google.com*

Mark Yoshitake  
*Product Manager, Cultural Institute*  
*Google*  
*Paris, France*  
*myoshitake@google.com*

Sertan Girgin  
*Software Engineer*  
*Google*  
*Paris, France*  
*sertan@google.com*

**Abstract**—The Google Cultural Institute Platform<sup>1</sup> is a large-scale system for ingesting, archiving, organizing, and interacting with digital assets of cultural material. This paper explains the components through which the platform contextualizes individual assets in order to enable storytelling. Contextualization is an inverse problem: given assets that are instances of cultural material, infer their precise context and use that as a way to support the storytelling process. The approach is based on three components: extraction, knowledge, and scale. Extraction is the inference of context from two sources of information: explicitly provided metadata, and automatically extracted features. Knowledge is the use of a large reference fact database for further contextualizing an asset based on its descriptors. And scale, achieved through global self-serve, enables massively expanded coverage of the knowledge database and crowdsourcing potential for metadata refinement. Together these components sustain a storytelling framework and a compelling user experience that has the potential to become the largest repository of cultural information and coherent narrative in history.

**Keywords**-knowledge management, semantic web, text analysis, image analysis

## I. INTRODUCTION

The Google Cultural Institute (CI) has launched a unified platform for storing, searching, and interpreting cultural material [1], [2]. This platform is based on digital assets in the form of images, video, and associated metadata, together with tools to manage their ingestion and artful placement within a broader narrative context via digital exhibits. While a primary focus of the system is to provide a compelling, engaging user experience through storytelling with cultural material, in this paper we focus on the challenge of positioning assets within their correct knowledge context in order to support the construction of compelling stories. Successfully meeting this challenge is important for the end users of the system, who avoid tricky and burdensome constraints at asset ingestion, and are rewarded with high level, organized tools for compelling, accurate storytelling over the appropriate assets.

Rather than considering ingested assets only as something to archive and retrieve [3], the CI platform considers assets as instances of knowledge represented in a large and growing

<sup>1</sup>google.com/culturalinstitute



Figure 1. President John F. Kennedy arriving at Love Field. Identifying the people in the photograph together with the date or location allows to link the asset with his assassination (LIFE Photo collection, Arthur Rickerby).

fact database (e.g., Freebase [4]). When an asset is correctly associated with knowledge it unlocks much more contextual information about the asset than may have been provided by the user during ingestion. For example, an image of John F. Kennedy may be difficult to interpret from knowing only that the image is of a person. But once the person is identified as Kennedy, with his life carefully and completely represented by the fact database, the scene can invoke a very specific context (Fig. 1). This is a kind of inverse problem - inferring the place in which an asset should sit within the broader context of the knowledge represented in a fact database. As such it is typically under-constrained, making it a real challenge to solve.

There are three sets of information to exploit in relating an asset to appropriate portions of the fact database. First is the asset itself, which embeds unstructured information that is challenging to extract through automatic means. Second is the metadata supplied by a user in the form of, for example, category tags and textual descriptions. And third is the separate fact database, which can guide inference because of the extensive set of fact and entity relationships it encodes. The fact database is useless, however, without solving a difficult matching problem, where assets must be associated with the correct portions of the fact database in

order to leverage its use.

The interplay between the kinds of information revolving around asset ingestion sets up an interesting set of challenges. Should the user be required to supply more or less direct metadata? How much can be automatically extracted? How reliable are the results from the automatic extraction? Once a composite set of descriptors is obtained, how hard is it to reliably match an asset to a precise place or set of places in the fact database? In this paper we describe three components in the Cultural Institute Platform that work together at the interplay of assets and their descriptors.

The first component is extraction. The platform provides a straightforward ingestion operation with the ability to specify descriptive metadata in a systematic schema. This metadata is valuable in understanding the asset to which it is attached. In addition, the platform applies a series of independent extraction algorithms to the asset in order to understand its content independently from the supplied metadata. This operation can be applied in parallel and can even be periodically re-executed as they are improved. The content extraction pipelines can be trained and tuned on large and independent image and video sets in order to improve performance. The descriptors from this process are collected and evaluated as to their reliability and together with the provided metadata represent all that is known about the asset without reference to an external source.

The second component is knowledge. Rather than building only a scalable storage repository for assets with associated metadata, the platform views cultural assets as instances of entities in a larger, independent graph of knowledge, such as that embodied in a large fact database. Thus ingested assets, as instances of information in the graph, are viewed as a means through which graph entities and relationships can be constructed, confirmed, and illustrated. It is arguable that cultural and historical entities can be reliably constructed only through such means, where large repositories dedicated to culture and history participate actively in its emergence. How to build the inference algorithms in order to correctly relate asset instances to entities is a major challenge.

The third component is scale through democratization. The CI platform embraces and embodies the desire to preserve, organize, and share our collective human cultural experience. Through the mechanism of self-serve and click-to-accept service agreements, the platform allows broad participation from groups who can ingest assets and build exhibits. Ingestion and exhibit tools are built to be accessible by average non-expert users, which can inspire researchers, teachers, independent artists, historians, and virtually any group or individual working with assets in the cultural sector to share them and grow the repository toward a truly large-scale system that is diverse and representative of almost all areas of human cultural experience.

## II. STORYTELLING

The CI platform is hierarchical, allowing curators to form compelling and nuanced narrative presentations from underlying ingested assets. This storytelling component is compelling for users who benefit from a narrative context for individual assets, and it sets the stage for the construction of context, in the form of artful narrative sequences, that relates sets of assets to each other. Once a digital exhibit is constructed and approved it can be viewed by a global audience, allowing users to engage with the story and at any time independently explore a particular asset.

For example, the digital exhibit “Bletchley Park: Home of the Codebreakers” (Fig. 2) shows the title slide of the story around the work at Bletchley Park. The entry slide shows the exhibit title, an overview bar along the bottom of the overall story with icons representing individual assets involved in the exhibit, and the first asset in the story: an image of Her Majesty the Queen at the public memorial for Bletchley Park veterans and its outstations, taken on 15 July 2011. The user can advance through the story, which includes writing and images and video positioned in sequence. Two assets from the exhibit (Fig. 3) illustrate how the nuanced context of a story can relate potentially disparate assets: on one hand Her Majesty the Queen, and on the other an image of the Enigma. Associating these assets makes perfect sense in the context of the narrative surrounding Bletchley and, more broadly, the events around the Second World War.

The storytelling platform is premised first on the ingestion of assets over which stories are constructed. But in order for it to be easy for stories over assets to be formed, it is crucial that relevant assets be organized and readily surfaced for the storyteller. With storytelling as a goal, the underlying components of extraction, knowledge, and scale must be well-managed and mutually supportive, which is a major challenge.

## III. EXTRACTION

Ingestion is the process of establishing an asset within the repository. An asset is a composite object consisting of content (e.g., digital image, video) and metadata. The ingestion process is centrally important for a number of reasons. From the users perspective, a simple and straightforward ingestion process facilitates rapid and efficient movement toward the higher-level goal of storytelling. In terms of the platform and its structure, ingestion is a crucial point where the asset can be contextualized and prepared for future requests, such as search and use within digital exhibits. And from the archivists perspective, ingestion is where the asset is definitively specified and registered within the system for safekeeping and future canonical retrieval.

In the CI platform, the metadata of an asset consists of a set of attributes, such as “title”, “description” and “creator”. Each attribute belongs to a schema (either the core schema that encapsulates common attributes or a partner-specific



Figure 2. A digital exhibit about Bletchley Park.

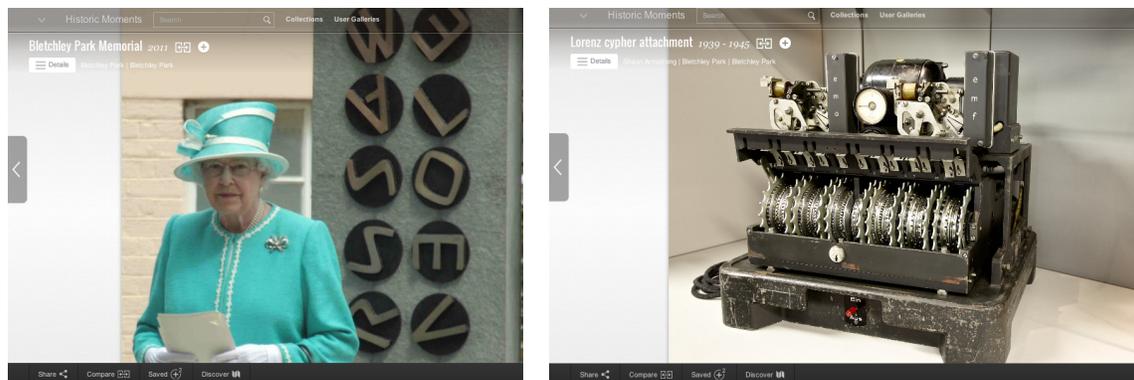


Figure 3. Examples of digital assets (images) used in the Bletchley Park exhibit.

one) and has a particular type. Attributes may be optional or repeatable (e.g. a title is required, description may or may not be present, and there can be multiple creators), and it is possible to refine them in a hierarchical manner by using suffixes, e.g. "creator.author", which enables encoding fine-grained semantic relationships.

Fig. 4 shows a schematic of the components of the ingestion process. At the end of the process the asset becomes a part of the repository. But before that is possible, the platform attempts to improve the ability to contextualize the asset using three strategies. These strategies and their results are coordinated through an "inference engine" that can assess the reliability of these attempts at extraction and asset contextualization.

The feature extraction process searches the content directly for indications of objects and their attributes based

on the content and on hints from the provided metadata. For example, "The Starry Night" is a "mostly blue" "post-impressionist" "oil painting" by "Vincent van Gogh" depicting a "night" "sky". Extracting the object references and attributes requires syntactic and semantic analysis of metadata, and computer vision, in particular image recognition and identification [5]. The repository can be consulted by the inference engine to determine if the asset is similar to content that has already been ingested. The result of these methods (provided metadata, extracted features, similarity search) produces a set of annotations ("entity annotations") for the asset. For a fully and completely specified asset, or one that is a perfect match with an already-ingested asset, the resulting annotations may be very specific and fully-formed at this point. However, it may also be the case that the annotation results are incomplete and less

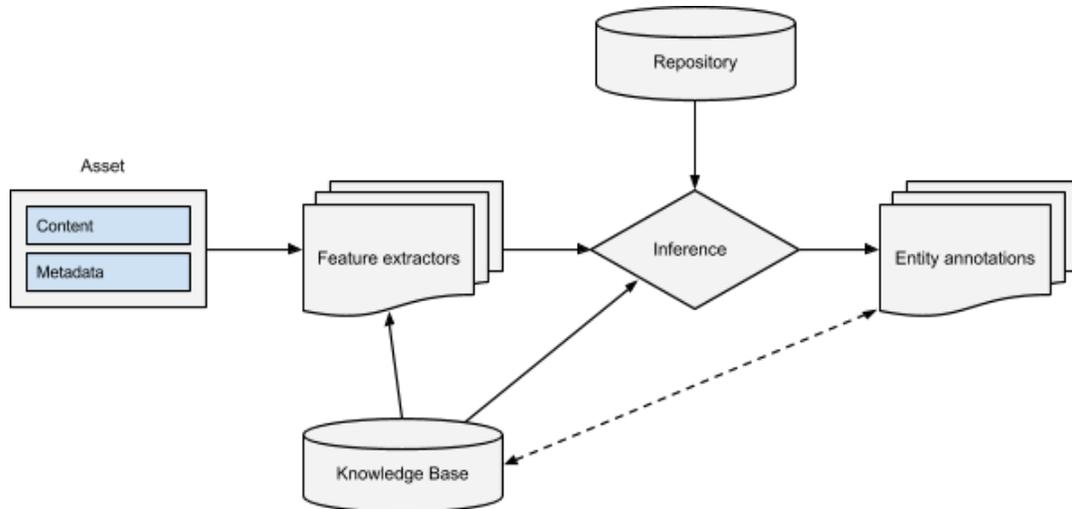


Figure 4. Components in the ingestion and extraction pipeline.

specific than desired. This is mostly due to the fact that the value of metadata attributes are usually available in free-text form only and not canonical. For example, the value of the location attribute may simply be "Paris" instead of explicit geographical coordinates. Without any additional information this leads to ambiguity as it may refer to the capital of France, a city in Texas, USA (and 15+ other states), or a hamlet in Denmark. The first interpretation is more likely but others are also possible; "Paris, Europe" is less ambiguous and "Paris, France" uniquely identifies the location. Furthermore, content-based recognition is also subject to noise and may not be 100% accurate. The platform includes a powerful follow-on step to improve the entity annotations, indicated by the dotted line in Fig. 4 between "entity annotations" and the knowledge base, which is more fully described in the next section. This link sets up a feedback loop between the other ingestion and extraction components, allowing for entity annotation refinement and reference over a very large fact database.

The components of this ingestion pipeline, driven by the inference engine, make it possible to derive accurate entity annotations rather than simply requiring the user to specify them. The accuracy of the derivation is dependent on reference to sources of information beyond what is initially specified in the form of the existing (presumably large) repository, the extraction algorithms (trained for accurate extraction on a large corpus of material), and the separate knowledge database. Fig. 5 shows examples of the extraction algorithms for annotations such as "wood", "chair" (Fig. 5a), "sea", "horizon", and "shore" (Fig. 5b).

#### IV. KNOWLEDGE

A large fact database, made of entities (person, place, thing) which have attributes and relationships, can be used to solve the challenging problem of understanding the precise

context of ingested assets. The CI Platform uses a fact database in order to associate ingested assets with broader contextual information than what may have been provided at ingestion time. This component is important as a support for storytelling because it makes it possible to produce assets during a search that are connected with the broader fact database instead of just the (potentially limited) metadata provided at ingestion. This means that many more assets can become discoverable to the storyteller for use in the process of constructing digital exhibits.

The platform uses two complementary approaches in order to associate an asset with facts in the database. First is the search for similar assets. When an ingested asset is determined to be similar to one that is already present in the repository, that association makes it possible to infer that the new asset can inherit all the context already attributed to the existing, similar asset. When this similarity association can be established it powerfully contextualizes the new asset. Fig. 4 shows the inference engine's connection to the existing repository - the similarity search gives the inference engine access to established annotations on similar assets.

Second is the attempt to infer a relationship between facts in the database and the asset metadata either that has been either provided or extracted. The inference algorithm uses the assets provided and extracted metadata and searches for matching support among the fact database. This is a challenging, imperfect process that depends on continued analysis in order to establish confidence thresholds that exclude spurious associations. Fig. 4 shows the inference engine's access to the knowledge base for the purpose of this matching process.

The current platform contains all the components shown in Fig. 4 and thus gives a unique opportunity for building and evaluating algorithms for extraction and for inference that

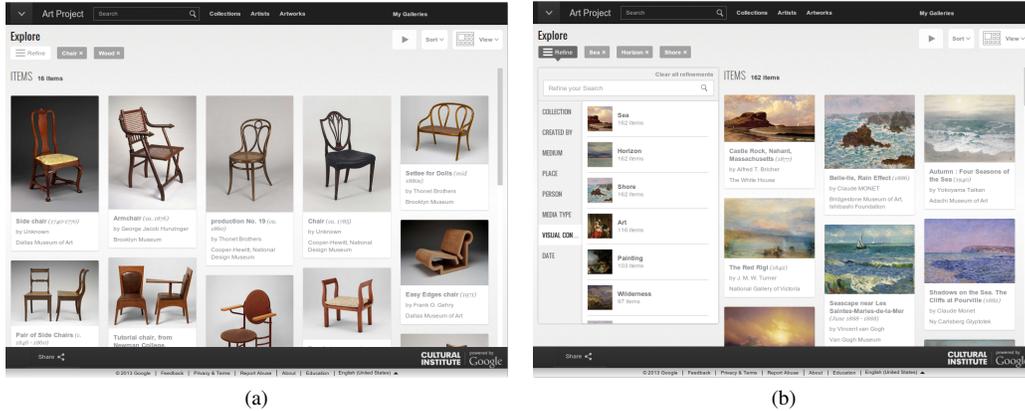


Figure 5. Examples of automatic extraction of visual content; objects annotated with (a) “wood” and “chair”, and (b) “sea”, “horizon” and “shore”.

are accurate. It is important to note that nuanced annotations are notoriously difficult to extract automatically, and it is still true that the best way to arrive at correct annotations is through human involvement. The valuable by-product of the CI platform, which integrates its canonical repository with storytelling through digital exhibits, is that it admits a human-driven crowdsourcing for entity annotations and corrections. Storytellers as well as those users immersed in stories told over annotated assets notice annotation gaps and can be encouraged to correct them. This aspect of the platform is likely to become an important part of the iterative inference engine as assets and stories grow to scale.

With crowdsourcing in mind, the platform provides a number of other attractive features that will engage users and attract them to the place where they can view, evaluate, and supply more information about assets. For example, users can construct their own exhibits with assets that they deem interesting, and those assets can be compared with a visual and interactive “compare tool” that makes side-by-side and layer-based comparisons simple and compelling. These exhibits can appear for other users in searches, inspiring new storylines and comparisons among ingested assets to continue appearing as the repository grows. These aspects of the platform lead toward the final crucial component: growing the number of assets and the number of digital exhibits to a very large scale.

## V. SCALE

The potential scale of a repository that collects and maintains instances of cultural information is enormous. Even with a very narrow definition of “cultural”, such as “art”, it is clear that the potential for large-scale information repositories is almost unbounded. In fact the definition of “culture” is very broad and it is important to build coverage over the diversity and entirety of human experience. While scale represents an engineering challenge, it brings many potential benefits for the CI platform.

First, a diverse and large-scale repository appeals to many

more users, allowing them to personally connect to the material and the stories. As the number of assets and stories grows, the number of people who connect with the material will also grow. As a canonical reference platform, growing to scale and covering more of the human experience allows a democratization of those experiences, giving voice to a diverse range of stories. And with precise organization of assets, instances can be produced for users who search for specific examples.

In a very practical way, growing the repository to scale will improve each of the components of the system as it is currently constructed. The ingestion and extraction phase will become more streamlined as the asset repository grows, systematically reducing the need for the specification of large amounts of metadata with each ingested asset. Scale also has a positive impact on the extraction algorithms, which can be trained to be more precise at extracting visual content over a larger and larger corpus of examples.

Perhaps most important is the construction of the knowledge database. As assets are added to a scaled-up repository, the knowledge database can be extended to cover more and more facts that form a sort of “entity backplate” for the human experience. As this database is grown, the entities it contains become richly exhibited as instances of assets and stories in the platform. It is currently a challenge to grow the fact database from scaled-up asset collection while preserving its accuracy and eliminating redundancies. But the fact database remains a powerful approach for contextualizing assets and for supplying context for storytelling and education.

The CI platform will enable scale-up through a democratized approach to ingestion and storytelling. Any user interested in using the platform is enabled through a simple terms-of-use agreement to provide content and supply digital exhibits. Increased usage through this worldwide availability has the potential to provide the kind of scale that can lead to radical improvements in content extraction, inference of complex annotations from content similarity, and automated

contextualization of content using the knowledge base.

## VI. CONCLUSION

The Cultural Institute platform is an integrated repository of digital content related to cultural themes: art, history, and artifacts from antiquity. The platform provides a unique way to contextualize the digital assets it ingests for the purpose of storytelling. By providing an architecture and algorithms that exploit the mutually supportive components of extraction, knowledge, and scale, the system can offer both canonical storage and an engaging user experience through digital exhibitions.

The challenge of requiring the content provider to supply exhaustive metadata around each and every ingested asset is made easier through automated matching algorithms that search for similar assets as well as visual content extraction to identify latent content that is not necessarily specified in the metadata. In addition, access to a large fact database enables an inference engine to contextualize assets within a large and growing set of entities, properties, and relationships. The architecture is built to scale through self-serve ingestion and individual users who are empowered to provide content. Scaling the number of assets, the size of the fact database, and the power of the visual content extraction algorithms then makes the system increasingly more accurate and refined in its organization and contextualization of its holdings.

This platform uniquely supports artful and compelling storytelling premised on the components of extraction, knowledge, and scale. The user experience is significantly improved for each of the user actions (content ingestion,

creation of digital exhibits, viewing content and understanding its interpretive framework) through these components. Ultimately it will be user inspiration that drives such a collection to become large, well-organized, and compelling to explore.

## ACKNOWLEDGEMENTS

The authors thank Victor Ribeiro and the CI Engineering team at Google - Paris for continued impact through implementation of the platform. W. B. Seales was a visiting researcher at the Google Cultural Institute during the 2012-13 academic year.

## REFERENCES

- [1] E. Pfanner, "Quietly, Google puts history online," *The New York Times*, 11 2011. [Online]. Available: <http://www.nytimes.com/2011/11/21/technology/quietly-google-puts-history-online.html>
- [2] "Google Cultural Institute launches new platform for cultural sector," Collections Trust, 10 2012. [Online]. Available: <http://www.collectionstrust.org.uk/googleinstitute>
- [3] B. R. Schatz and H. Chen, "Building large-scale digital libraries," *IEEE Computer*, vol. 29, no. 5, pp. 22–26, 1996.
- [4] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: a collaboratively created graph database for structuring human knowledge," in *SIGMOD '08: Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. New York, NY, USA: ACM, 2008, pp. 1247–1250.
- [5] D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques," *Pattern Recognition*, vol. 45, no. 1, pp. 346–362, 2012.