Cultural Heritage Spatial Data Infrastructures (SDI) - Unlocking the Potential of our Cultural Landscape Data

Anthony J CORNS\textsuperscript{a,1}, and Robert SHAW\textsuperscript{a}
\textsuperscript{a} The Discovery Programme, Ireland

Abstract. The creation and use of digital spatial data within cultural heritage studies has increased over the past decade. The commissioning of high quality LiDAR, photogrammetric and geophysical datasets has brought about a better understanding of how human activity has shaped our landscape. However, the ability to realize the full potential of this cultural heritage data through re-use and sharing with external organizations and the wider research community is limited by technological, semantic and organizational barriers. Capturing and processing primary spatial data is expensive and any failure to fully exploit the investment in such cultural heritage data could result in reluctance to commission future projects. The development of open standards and technologies has enabled the creation of spatial data infrastructures (SDI) that allow users to discover, evaluate and use spatial datasets, and incorporate them into their own research. Much of the effort to date has gone into harmonizing environmental and security related spatial data. Cultural heritage information experts should be at the forefront adopting and adapting many of the standards established within the wider research community. This paper explores the requirements and research necessary for the development of a spatial data infrastructure (SDI) for the sharing and re-use of cultural heritage data, allowing for the efficient understanding and management of our shared cultural heritage.

Keywords. SDI, GIS, remote sensing, LiDAR, geophysics, e-infrastructures, photogrammetry

Introduction

The Discovery Programme, in common with many organizations involved in archaeological research or cultural heritage management is generating increasing volumes of high quality spatial datasets in the course of its research projects. It is now commonplace to carry out geophysical surveys, generate orthoimagery and DEM’s from photogrammetry, or commission high resolution LiDAR surveys, all rich resources which are used to further our understanding of the archaeological record.

In many cases funding for this data collection comes – directly or indirectly – from public sources, through national or EU funding, but with no requirement or mechanism to make the collected datasets available to other potential users. Once research has been completed and published, the underlying spatial datasets often remain hidden from, and unavailable to, the wider research community, thus failing to realize their full potential.

A group of like-minded archaeological researchers in Ireland recognized this problem and expressed the common desire to open up and share the resources they independently held. In 2008, through an Irish National Strategic Archaeological Research (INSTAR) grant, a feasibility study was undertaken, coordinated by the Discovery Programme to examine the mechanisms and technologies which would allow the sharing and re-use of our spatial landscape data by the wider community. Called the Spatial Heritage & Archaeological Research Environment I.T. (SHARE IT) project this one year study examined many of the issues related to archiving, accessing and sharing data, and concluded that Spatial Data Infrastructures (SDI), specifically for archaeological and heritage landscape data would provide a possible solution. This paper examines the reasons why SDI appear to offer the

\textsuperscript{1} Corresponding Author: Anthony Corns, The Discovery Programme, 63 Merrion Square, Dublin 2, Ireland; E-mail: anthony@discoveryprogramme.ie
way forward, and looks at the exciting possibilities that a newly announced EU Culture 2007 project, Archaeolandscapes, presents to put into practice an EU wide cultural heritage SDI.

1. Current obstacles to accessing and sharing data

As mentioned in the introduction much financial and professional effort has been invested in the collection and analysis of spatial data for archaeology and cultural heritage by state agencies, academic researchers and even the commercial sector. The gathering, cleaning and processing of these datasets has been carried out at great financial expense and personal effort often for specific research projects.

1.1. Data silos

Once datasets have been used for their original research purpose they are often stored in an unstructured way on the source organization’s server, or back-up tapes. This may be deemed an acceptable form of storage from the source organization point of view, but even then it is vulnerable to a number of serious archiving problems such as obsolete formats and software, corruption, loss of knowledge. Furthermore, the datasets are being hidden from and have become inaccessible to other potential users, eliminating any possible knowledge transfer. The result of this approach is the creation of fragmented and discrete data silos – vaults of information containing non-standardized data types. This approach essentially locks away data from the wider research community, and as a result the risk exists of replication of effort in creating, processing or cleaning data by other researchers.

There are also several barriers to the sharing of data which can be broadly grouped into the following three categories.

1.2. Institutional barriers

The intensely competitive environment that exists between research institutions can result in understandably protectionist attitudes, and a reluctance to share resources. However, if creators of data look beyond their own field of study, the data may be of great value in different field of scientific study. A LiDAR dataset collected for landscape archaeology could prove invaluable to a geomorphologist, and vice versa.

1.3. Geographic barriers

These exist due to the fact that, internationally, there are very different strategies for accessing data and there are currently no agreed mechanisms to explore and share data across geographic boundaries. There is generally less exposure to, and awareness of international datasets beyond individual researchers’ special interests, and researchers rarely establish formal cross-border relationships unless structured within an EU or international scheme.

1.4. Semantic barriers

Culturally sites and features may be described and represented in very different ways. National archaeological inventories will use classifications, structures and terminology appropriate to the specific country. Harmonizing the inventories of more than one country is highly complex, mapping across cultural themes and requires huge effort.
2. The framework for a solution

Over the past decade several separate developments have taken place that linked together might create the platform to solve many of the problems previously outlined.

2.1. International standards

There has been the establishment and adoption of standards for storing, describing, and accessing spatial data by a number of agencies, most notably implemented by the International Standards Organization, through ISO 19115\(^2\). This effectively provides a rulebook to data creators ensuring datasets are formatted in a standardized way, appropriate for sharing and re-use by others.

Central to the success of this is the provision of standardized metadata to describe the spatial data, vital to enable discovery of the resource in the future.

2.2. Open Geospatial Consortium (OGC)

The OGC has done invaluable work\(^3\) creating voluntary consensus at all levels of the Geographic Information (GI) community to encourage and enable interoperable solutions. The result of this has been to bring GI into the public realm, and helped to geo-enable the web, both significant factors that promote the open sharing of spatial data.

2.3. Service Oriented Architecture (SOA)

The development and adoption of SOA has facilitated the exchange of data using a well-defined, shared format that enables users to reuse and combine services reliably for their own purposes.

Archaeology and cultural heritage should be adopting all three approaches to spatial data and already has done so in some notable cases such as the ARENA2 project\(^4\).

3. Spatial data infrastructure

SDI is the integrated embodiment of the standardizations and technological approaches advocated in the previous section, all vital components that enable the harmonization of geospatial data. The SHARE-IT research project concluded that the solution for Irish cultural heritage lay in the establishment of SDI which would enable the discovery, evaluation and reuse of heritage and archaeological spatial data\(^5\).

3.1. Concepts of SDI

To help understand the concept of SDI a useful definition is found in the SDI ‘cookbook’:

“…the relevant base collection of the technologies, policies and institutional arrangements that facilitate the availability of access to spatial data”\(^6\) Figure 1 shows a schematic diagram of the basic concepts of SDI.

- The author creates a spatial dataset or map using the convention approach, such as a desktop Geographical Information System (GIS), but generates two extra key components.
- Firstly the author creates and publishes a metadata record that references the new spatial dataset, conforming to internationally recognized standards.
- The author then publishes the spatial data or map via web mapping services utilizing the interoperability standards defined by the OGC.
- A consumer wishing to discover this dataset will utilize a metadata catalogue or geo-portal and find appropriate data using free or agreed theme search criteria.
Once data is discovered its web mapping service is exposed and the user can consume it within a provided web application or on their desktop software.

**Figure 1.** Schematic diagram of the components of SDI.

There are important additional elements which should be considered along with the development of SDI. The organizational structure and commitment needs to be in place to keep the cycle of knowledge reuse in operation and a trusted digital repository is necessary, ideally based on the principles of the OAIS\(^7\). For the SHARE-IT project Fedora Commons\(^8\) was considered to be a solution.

### 3.2. Current SDI developments

Looking beyond the narrow confines of archaeology and cultural heritage considerable advances have been made globally in the development of SDI in recent years. The EU has two major schemes in place relating to SDI development.

- The European Earth observation programme, Global Monitoring for Environment and Security (GMES)\(^9\) consists of a complex set of systems collecting environmental data from multiple sources and providing it to policy makers online – hence it has an important SDI component.
- A partner scheme, the Spatial Information in the European Community (INSPIRE) aims to provide the standards and policies that will enable the integration of these services.

Both these programmes resulted in significant research funding through a number of FP6 projects including Humboldt\(^10\), eSDI-NET\(^11\) and the BOSS4GMES\(^12\) project which researched all aspects of SDI development from data harmonization to the legal aspects of data sharing.

SDI research and development is not restricted to the EU with initiatives in the USA at both federal and institutional level. Commendable projects include the NASA GeoBrain\(^13\) and Geospatial One Stop Portal\(^14\) projects.

Looking at these examples from the EU and USA it is apparent that the focus of research and development is on areas relating to security and the environment. Cultural heritage or archaeology are not currently included in any of these projects, hence the desire to look at the possibility of a designing SDI specifically for cultural heritage. The Irish funding for SDI research, the SHARE-IT project, ended in December 2008, but the Discovery Programme is now a member of an EU Culture2007-2013 project called ArchaeoLandscapes\(^15\) which has received funding for the next 5 years. Within the wider objects of this project it is hoped to move forward on filling the cultural heritage SDI gap.

### 4. SDI design for ArcheoLandscapes

Within the wider objective of the project the aim is to construct a system for the delivery of aerial archaeology and remote-sensing data to the general public, education and academic research com-
munities. In preparation for this objective a system diagram has been generated to identify the major components and how they will interact, (see Figure 2).

The following sections briefly add some detail to the concepts and interactions of the main elements of the system diagram.

Figure 2. Schematic diagram of the components and interactions of the proposed ArchaeoLandscapes SDI.

4.1. Data archive

The SHARE-IT project highlighted the importance of creating an archive for spatial data conforming to the best international practice. For this project two digital data repositories are proposed, one to deal with the truly spatial datasets, such as orthoimagery and geophysical data, and another to deal with the non-spatial data such as non-rectified oblique aerial photography.

A geodatabase approach is proposed for the truly spatial data, with Inspire compliant metadata which will be accessed via a spatial web server. The project will assess both open source and proprietary software solutions, documenting the processes to enable other partners to establish conforming archives.

The Fedora Commons open archive model is seen as the best solution to the storage and access of the non-spatial datasets and associated metadata. Research is planned into the potential to migrate these datasets into spatial data through georectification.

4.2. Geoportal

This is the vital component within the system, the development of a single website which acts as the hub for the user’s spatial requirements, providing the tools to search, discover, access and consume spatial data.

The concept of the geoportal has been implemented to great effect to deliver national infrastructural and environmental spatial data, examples such as the UK Go-Geo and US government’s geodata websites. Geoportals such as these enable browsing of data by theme, provider, temporal or spatial relevance, all of which is derived from the standardized metadata. By inspecting the full
metadata record the user can assess if the data is fit-for-purpose, and then consume it as a live web service from the geoserver.

However one of the great strengths of the geoportal is its ability to harvest metadata records from remote archives, such as those of our ArchaeoLandscape partners, and construct a metadata catalogue. The result would be a geoportal providing a single access point for European archaeological spatial data. This widening of access to data will inevitably raise important and legitimate concerns over the issue of access management and copyright control but this could be dealt with through the implementation of the Creative Commons.

4.3. User applications

Using web services as the delivery mechanism for spatial datasets has the significant advantage that they can be consumed by a range of devices and software applications providing a user experience appropriate to the consumer’s requirement.

Thin-client web applications such as Google Earth or Microsoft Bing Maps provide a software environment which is both free and familiar to a rapidly growing section of the public.

Alternatively, the more experienced GI user can consume datasets via web services from remote servers to their standard desktop GIS system such as ArcGIS. The advantages of this are that the user retains access to their full suite of processing and analysis tools, and accesses data of reliable quality and currency without storing a copy of the data locally.

Another increasingly common approach is to custom design a tailored web mapping application. The development of software such as Adobe Flex and Microsoft Silverlight provides a simple, fast way to develop what are known as rich internet applications. These web applications, with GIS functionality, place an emphasis on the quality of content and have the potential to significantly increase the user community. In the case of ArchaeoLandscapes this would provide the tools to promote cultural heritage data effectively to a broader community thus generate huge educational benefits, both core objectives of the project.

4.4. Content integration

ArchaeoLandscapes will generate other non-spatial content such as supporting documentation and media that will improve the understanding and awareness of the primary spatial datasets. Integrating this content into the system via content management system (CMS) such as Drupal or Joomla will present the user with a more complete and seamless user experience. An example of this could be the integration of teaching material prepared by a partner organization being linked to appropriate live spatial datasets to create an integrated lesson which students around Europe could access and use at their desktop.

5. Conclusions

Preparation for the ArcheoLandscapes project has been built on the solid understanding of the benefits of, and need for, SDI for cultural heritage which was the conclusion of the SHARE-IT project. Such a development would realize the true value of datasets currently locked away in disparate servers, with the immense benefits to education and research which would follow.

Technological advances will inevitably impact on the components and interactions envisaged at this stage, so it important that the project design is flexible. However the core principle, the benefits in opening access and sharing our data, should remain.
References

3 Interoperability & Open Architectures: An Analysis of Existing Standardisation Processes & Procedures, Martin Klopf (Editor),
   http://portal.opengeospatial.org/files/?artifact_id=10594
http://ads.ahds.ac.uk/arean/ARENA2.html
5 http://www.share-it.ie/index.html
6 Nebert, D. ed., 2004. Developing Spatial Data Infrastructures: The SDI Cookbook. Available at:
16 http://http://www.gogeo.ac.uk/