The current issue and full text archive of this journal is available at
www.emeraldinsight.com/0737-8831.htm

Received 4 July 2013
Revised 12 August 2013
30 September 2013
Accepted 12 October 2013

REGULAR PAPER

The Albert Einstein archives digitization project: opening hidden treasures

Dalia Mendelsson and Edith Falk
Library Authority, Hebrew University of Jerusalem, Jerusalem, Israel, and
Amalya L. Oliver
Department of Sociology and Anthropology, Hebrew University of Jerusalem, Jerusalem, Israel

Abstract

Purpose – The purpose of this paper is to present the organizational and technological processes and strategic choices that led to the successful digitization project of the Albert Einstein Archives.
Design/methodology/approach – This is a case study of the major challenges that were associated with the project. These include: the integration of the archives in the academic environment; the management of a project of such magnitude within the university organization and between different stakeholders and the technological aspects of the project and user experience.
Findings – A digitization project requires not only the archival staff expertise but also information specialists, IT staff, analysts and usually the digitization staff for processing the archival material. Finding the common language between all the professionals involved as well as building a good strategic plan are the keys to a successful project.
Research limitations/implications – The planning and implementation of such a project requires a significant budget, manpower project management, hardware, software and intra- and inter-organizational cooperation and coordination.
Originality/value – The phenomenon of digitizing unique and exclusive archival data by universities is becoming an innovative contribution of hidden goods to the public at large. This paper offers strategic insights for the planning of similar digitizing projects, particularly in an academic environment.

Keywords Information technology, Archives, Digitization, Project planning, Academic environment, Memory institutions, LAM, Albert Einstein archives

Paper type Case study

Introduction

During the past decades, universities are considering opening their archival treasures to the public. This challenge involves the planning and implementation of a digitization process; a process that requires technology know-how, a significant budget, manpower planning and developing IT capabilities including hardware and software. In addition this process also requires organizational cooperation and interaction within the academic institution and other stakeholders.

On 19 March 2012 the Albert Einstein Archives (AEAs) digitization project was launched at the Hebrew University of Jerusalem. The official web site of Albert Einstein, (http://alberteinstein.info) including the Einstein Archives received more than 21 million

The authors would like to acknowledge The Polonsky Foundation; Professor Hanoch Gutfreund; The Einstein Papers Project, at the California Institute of Technology; The Albert Einstein Archive, The Library Authority, The Hebrew University of Jerusalem and Ardon Bar-Hama.

Library Hi Tech
Vol. 32 No. 2, 2014
pp. 318-335

0737-8831
DOI 10.1108/LHT-07-2013-0084

hits on the first five days and over 650,000 unique visitors from 172 countries. The press coverage delivered over 420 online articles and broadcast media items.

This article presents the implementation of the AEAs digitization project emphasizing the following aspects of the digitization process:

- integration of the archives in the academic environment;
- managing a project of such magnitude within the university organization; and
- the technical aspect of the project and user experience.

Background

About the AEAs collection

Albert Einstein (1879-1955), one of the most influential personalities of the twentieth century, bequest all his intellectual property to the Hebrew University of Jerusalem. This includes: personal letters, family correspondence, scientific and non-scientific writings, prizes, medals, diplomas and photos. Albert Einstein was a poor record keeper and had to rely on Helen Dukas, his assistant for three decades, to organize his vast and diverse paper work. After the scientist's death in 1955, Dukas and Einstein's close friend, Dr Otto Nathan spent a quarter of a century organizing the papers and acquiring additional material. The archive was transferred from Princeton University to the Hebrew University in 1982.

The AEA at Hebrew University holds the largest collection of Einstein's original manuscripts in the world. The intellectual property was carefully catalogued, first manually by Helen Dukas and then in a propriety cataloguing system developed at Caltech, California by the Einstein Papers Project (EPP) staff. EPP edits the Collected Papers of Albert Einstein (CPAE), published by Princeton University Press (PUP). The cataloguing and the enrichment of the database is, to date, a joint project of the EPP and AEA.

After the inclusion of the AEA in the library system of Hebrew University in 2008, a digitization project was planned and started in 2011 – as the best way to preserve the archival material and present it to all audiences in an enhanced format.

This project signifies the commitment of the Hebrew University of Jerusalem to open the privately owned exclusive archive to the general public as declared by Prof Menachem Ben Sasson: "Knowledge is not about hiding. It's about openness[1]".

The challenge of integrating digital archives in the academic environment

Libraries – Archives – Museums

The inclusion of the AEAs in the library system of the Hebrew University is not surprising in an era of digital convergence when distinctions between archives, libraries and museums become blurred and unimportant (Kirchhoff *et al.*, 2008). The topic of the "digital convergence" of the LAMs[2] – as they became known – is not new. Rayward (1998), for example, examined early on how changes from physical to digital media affect the traditional distinctions between information organizations in his article on:

> […] electronic information and the functional integration of libraries, archives, and museums.

The idea of looking together at libraries, archives and museums has also served as the theme for several conferences, such as "Libraries, Archives, & Museums – Three-Ring Circus, One Big Show?" in 2005 July RLG Members Forum (2005) and "Libraries, Archives,

and Museums in the Twenty-First Century: Intersecting Missions, Converging Futures?" in 2006. More recently, OCLC has published a thorough and interesting study, "Beyond the Silos of the LAMs: Collaboration among Libraries, Archives and Museums" (Zorich *et al.*, 2008). The centrality of the subject is further emphasized by the decision taken by three very important journals in the three disciplines – *Library Quarterly*, *Archival Science* and *Museum Management and Curatorship* – to each publish a special issue:

> […] exploring the shared information needs and challenges facing libraries, archives, and museums in the information age; the overlapping educational goals of library and information science, archival studies, and museum studies programs; and areas of convergence for educators and professionals working to meet user needs in libraries, archives, and museums (Marty, 2008).

The blog by OCLC staff, "Hanging Together; the hangout spot for libraries, archives, and museums[3]" became a popular source for professionals in all three disciplines. The stage was now set for the birth of a new institution, the "memory institution", a birth made possible by the massive move to the digital environment, where the physical issue is much less crucial. This emphasizes the importance of accessibility and brings the digitized materials to the fore.

Lorcan Dempsey (2000), in a report prepared for the European Commission, presents a research framework for libraries, archives and museums in which he defines the concept:

> Archives, libraries and museums are memory institutions: they organize the European cultural and intellectual record. Their collections contain the memory of peoples, communities, institutions and individuals, the scientific and cultural heritage, and the products throughout time of our imagination, craft and learning. They join us to our ancestors and are our legacy to future generations. They are used by the child, the scholar, and the citizen, by the business person, the tourist and the learner. These in turn are creating the heritage of the future.

Traditionally, the three types of institution have quite different ways of organizing, documenting and preserving their collections. However, with the emergence of digitization as the main tool of preservation and presentation of all cultural artifacts, the digital memory institution was born. This new institution does not compete with the existing ones. On the contrary, says Dempsey – the digital tools increases the visibility of the traditional institutions.

A few years after the preparation of the Dempsey report, in 2005, French president Jacques Chirac, together with five other heads of state, wrote a letter to the president of the European Commission. The letter recommended the creation of a European virtual library[4]. This was the catalyst for the creation of Europeana, a web portal allowing free access to the cultural wealth of Europe's museums, libraries, archives and audiovisual collections[5].

To date, the portal includes metadata for over 29 million items. The National Library of Israel is one of the data providers. It has contributed so far a modest 3,846 items to the collection. The portal is one of the finest examples of the way the LAMs can be included in one database.

The convergence of the library and the archive in the digital age became a focal point of the theoretical understanding of these institutions. Inspired maybe by the French philosopher Jacques Derrida's (1995) "Archive Fever"[6], scholars such as Marlene Manoff (2006) and Michael J. Paulus (2011) look at the concept of the library and the archive, at the basic change that occurs in the nature of the book when it becomes an e-book and at the nature of the record when it is digitized. The distinction

between book and record is then blurred and so is the distinction between the library – the historical archive of the book – and the archives – the repository which documents a specific activity. Archives and libraries are then merged in a new archival cycle. Paulus (2011) states that by:

> […] linking more theoretical understandings of the archive, as an initial act or historical manifestation of inscription, with the more technical process of preserving such traces, academic librarians and archivists have the opportunity to build on the recognized value of the library as an archive, to position the library as a site of creation, to confront the reality of digital 'archives in the wild,' and to re-conceptualize their roles within the archival lifecycle.

Integrating the AEAs within the Hebrew University library authority

The Hebrew University libraries started to deal with archives only in 2008 – a fact perhaps due to their history: with the opening of the Hebrew University in 1925, the National Library, which was founded in 1892 as a world centre for the preservation of Jewish thought and culture, became the Jewish and National University Library (JNUL) and an integral part of the Hebrew University. With the growth of the university, several subject libraries developed in parallel and became the main teaching and research libraries in most disciplines.

Over the years, the status of the JNUL was evaluated by several committees and in 2007, the Israeli Knesset enacted the National Library Law. On the 1st of January 2011, The National Library of Israel was officially separated from the Hebrew University and became an independent institution. However, it remained the central research library of the Hebrew University of Jerusalem in the subjects of Jewish Studies, Islamic and Middle-Eastern Studies and in the General Humanities.

In parallel, the Hebrew University libraries continued to grow and develop. In 2003, they underwent a general reorganization which was mainly intended to put all libraries – at the time 15 libraries on four campuses – under the umbrella of one administrative and professional entity which was designated as the Hebrew University Library Authority.

During the existence of the JNUL, all archives, manuscripts, most special collections that belonged to the Hebrew University were housed there. These included the AEAs and private library which were willed by the scientist to the Hebrew University of Jerusalem. With the separation between the two institutions, almost all the material belonging to the University was left to the National Library on permanent loan. This did not include any of the Albert Einstein material that was transferred to the Hebrew University Library Authority in January 2008.

One archive and eight libraries

The integration of the Einstein Archives into the library world was not obvious. The importance of owning such a treasure as part of the intellectual property of the university was well recognized – today still, almost 60 years after the death of the scientist, the name Einstein instantly brings to mind the connotation of genius, science, innovation, etc. Moreover, by the time the Einstein Archives were integrated into the library system, it was clear that the unique holdings of an institution is what makes it different from its peers[7] and the potential value of these specific archives was quite stunning: it must be accessible to all audiences, with the help of state-of-the-art Web 2.0 applications and next generation discovery tools.

However, the actual technique of making the best of this wonderful collection had to be dealt with. The Library Authority had to prepare a suitable, secure and

climate-controlled facility for the archives. After the collection which includes close to 80,000 items was transferred to its new quarters, the Library Authority had to find a way to maximize the potential of its new member. The concept of digitization was quite obvious. Already in 2001, Judith Panich (2001) noted that 81 per cent of ARL[8] members were conducting digitization projects of their special collections. The considerable improvements experienced since in software, hardware, metadata standards as well as the lower costs involved made the solution very clear: the whole collection must be scanned in high quality and the existing software must be replaced by a solid and well-known data management system.

Many articles have been written about digitization, but:

> […] how libraries will organize to provide long-term support for the creation of online resources in special collections is open to question (Sutton, 2004).

The first step was to secure funding for the project. Following the international innovative initiative of The Polonsky Foundation[9], a proposal for digitization of the Einstein Archives was prepared and presented to the Foundation which accepted it. The following step was to hire a project manager who would be able to complete the mission, and to put in place a strategic plan, involving the various units both inside and outside Hebrew University that would collaborate to achieve a successful outcome.

Collaboration to success

Multi-unit organizational or inter-organizational collaborative projects are leading sources for organizational innovation (Oliver, 2009; Powell *et al.*, 2005). In such collaborations, the most valuable knowledge, expertise and capabilities are located between organizational units and specialized actors, thus the process leads to valuable integration of these capabilities and knowledge. In cases where issues of property and intellectual rights are involved and different organizational actors have different interests and motives, there is a need to develop a working collaborative system. This system must take into account the need for conflict resolution processes in order to achieve the common goals and final project integration. This collaborative model is also characterized in the model of "open innovation" presented by Chesbrough (2003) and will be further discussed below.

This task of establishing a successful intra- and inter-organizational model requires a fulfilment of a set of critical steps. We will highlight the elements and explain their importance.

Identifying the central coordinative units

Before starting a collaborative process, it is important to identify one coordinative unit. This unit needs to include a dedicated organizational agent who is clear about the goal of the process and has both the technical and the inter-personal capabilities to coordinate large-scale projects. It is possible to have two project leaders – one who is a technical expert and the other who has the capabilities to identify the needed capabilities and knowledge for the project, generate enthusiasm and commitment, integrate the activities of the group members and anticipate conflicts and minimize their potential impact on the success of the project in advance.

Understanding the project goals and clarifying expectations of group members

Once identified, the central coordinative unit needs to build clarity on two levels: management and legitimacy; and analytical specifications. The first level needs to

insure support of the goals of the project by the high management of the organization. Without such support, large-scale coordinative projects will have limited legitimacy and will find it hard to attract other members to the process. For establishing inter-organizational collaborations, such activities need to involve central members of all the organizations participating in the project. These members need to be able to independently approve the needs raised by the project activities in real time without generating delays that can constrain the project.

The second element has to focus on the analytical clarification of the project goal and the major components needed for achieving successful outcomes.

Mapping the big picture

Large-scale projects are usually not conducted within one institutional roof. There is a need for experts in different domains that ought to collaborate across organizational boundaries (Oliver, 2009) and add their unique capabilities to the overall project. For this purpose, at this stage of a project, there is a need to map the large picture associated with the project. First, the main actors have to be listed along with their expected contribution to the project. It is important to identify a matrix of role of each expert, the time frame and the contingencies of their contribution. It is important to identify the role of each member in advance but at times, each contribution is contingent upon the accomplishment of the previous one. In the case of the Einstein project, it was necessary to obtain the rights to open the digitized manuscripts to the public from PUP, based on an existing agreement between the publisher and Hebrew University, before designing the public interface. Other tasks could be accomplished in parallel. For example, it was possible to work on the metadata migration, digitization of the images and the support of expected heavy traffic accessing the web site.

Clarifying tasks and roles and identifying "language" gaps

At an early stage of the digitization project, it became apparent that there is a need to clarify both the expectations and tasks of each expert and team. This clarification stage is most important in every collaborative project. We realized that the working "language" used by the different teams was tacit to each team based on the working culture they have developed over the years. Yet assumptions made on the statements were at times different for each team. The distinction between tacit and explicit knowledge is well explained by Nonaka (1994) who argued that:

> […] organizational knowledge is created through a continuous dialogue between tacit and explicit knowledge […]. It is argued that while new knowledge is developed by individuals, organizations play a critical role in articulating and amplifying that knowledge.

This distinction between tacit and explicit knowledge is important since members of organizations are not always aware of what they know and how much they know until they need to exchange some of their knowledge-base with others.

The interactive processes require a translation of tacit to explicit knowledge, a process that eventually accommodates the transfer of knowledge between group members. The digitization project brought together several units with their own expertise: the Library Authority and the Computer Authority at the Hebrew University and the staff from the EPP at Caltech, California Institute of Technology. The interaction between all the units for such a large-scale project in a short period of time required a "common language" in order to translate the project's goals and activities into practical tasks and to achieve the vision for this project to its success.

Running integrative and collaborative sessions

Even after the expectations, understanding and usage of working language is shared between different teams within and between organizations and clarification is achieved, the collaborative task is not over. At this stage, it is important to establish periodical working sessions where the progress made by all teams is reported, and new problems are examined. The ongoing collaborative exchanges are important for the purpose of establishing a stronger collaborative culture and increase the commitment of teams and experts to the joint project.

The main aspect of these on-going meetings is that, after establishing working relations, these have to be retained and improved. This is true especially since, in such emerging unique projects, the learning curve of all contributing participants is quite demanding. Routines need to be established and feedback loops need to be checked. Such closure systems are hard to achieve even within organizational close teams, not to mention experts operating as independent contractors. In this context, nothing should be assumed or be taken for granted and every detail and task has to be checked and examined. These exchange meetings should always keep in the background the joint goal of achieving a successful project based on joint interests and commitment of the participating teams and experts.

Resource management

The distribution of resources needed for such reputable projects have to be well managed. First, due to the high level of uncertainty of such projects with regards to needed resources coupled with a high level of precision needed for a "made to perfection" end result, resources need to have an additional slack. This means that in order to avoid unanticipated or unforeseen resource problems, the initial resource allocation of such projects have to include a significant slack level. The Einstein project benefited from a generous donation, and this should apply to any other projects of this calibre.

Temporal arrangements are also an important resource. A project needs to include time slack to allow for unexpected and unforeseen complexities and for a deeper process of learning. If, due to the flexibility given to the temporal dimension, the project reaches its goals before the designed deadline, there is plenty of time left for checking and reviewing the end product and for consulting with external auditors. Such projects can benefit from comments given by potential users and other experts in the notion of "open innovation" (Chesbrough, 2003). The idea of "open innovation" focuses on the importance of using a broad range of knowledge sources for an organization's innovation and inventive activities. This includes customers, competitors, other academics or administrators, and other organizations in unrelated industries. All these knowledge sources are used simultaneously using creative methods to exploit the organization's knowledge base and resources (West and Gallagher, 2006).

Monitoring potential conflicts of interests – establishing conflict avoidance and conflict resolution mechanisms

Conflicts in collaborative intra- and inter-organizational projects of the magnitude of the Einstein project are highly potential. This is mainly because there is competition for acknowledgement and credits (Oliver, 2004). In addition, at times, project members may feel deprived of the resources others are receiving or in time conflict with their other tasks within their organization. It is important to anticipate such conflicts and try to avoid them.

The challenge of presenting Einstein to the world

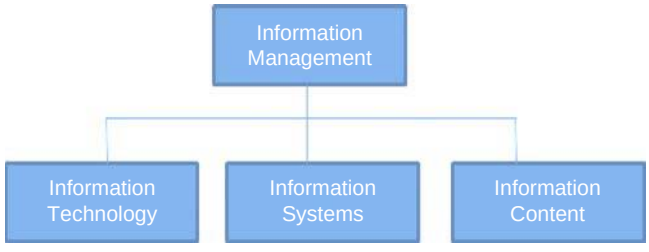
During the past decade, many libraries and archives worked on making hidden treasures available online to the public with the support of philanthropic funds, thus democratizing the knowledge which became free and open via the internet. Using the technological advances for supporting, viewing and preserving non-born digital data, these institutions started many digitization projects. The AEAs followed the trend – the Hebrew University Library Authority, supported by the Polonsky Foundation, started an ambitious project including the high-quality digitization of the whole archive as well as the migration to a new automated system. In this way, by transforming the AEAs into digital archives, the end user is offered the opportunity to interact with the material in a non-linear fashion as was indicated by Bolick (2006) – and to discover the works of the scientist as well as his personal and public life.

The project plan

1. *Overview.* The archival database includes the over 80,000 records catalogue of all the documents held in the Einstein Archives. It includes more than 40,000 personal papers of Albert Einstein, and more than 30,000 additional Einstein and Einstein-related documents discovered by the editors of "The Collected Papers of Albert Einstein" since the 1980s and by the archivists and curators of the AEAs at the Hebrew University.

Steven Buchanan (2010) describes the importance of preparing a strategic plan in order to design digital library services and affirms that the key to any successful project starts with such a plan. Understanding the archival material on the one hand as cultural heritage[10] and on the other hand as science information[11] (Chowdhury, 2010) was the base for developing the strategy of this project. This strategy was based on the holistic relationships and analysis of the following components: information management, information technology and information content as illustrated by Figure 1 (Buchanan, 2010).

Steven Buchanan mentions various models of strategic plans. One of these models indicates the following actions: vision, analysis, goals, strategic choices, implementation and courses of action[12]. In the case of the AEAs digitization project, the main idea was to present the work and life of the intellectual giant to the public as well as to preserve it for future generations. In order to achieve that vision, it was necessary to analyse the contents of the archive and to understand how to present the metadata and the digital objects in an interesting way, while adding as much useful and new perspectives as possible. It is on this background that the goals for this project such as online accessibility to the original material, navigation and discovery process of the archival material in order to preserve the universal asset and engage all audiences, led to strategic choices. These strategic choices will be presented in the following sections.



Source: Buchanan (2010)

Figure 1.
Information strategy components

2. Digitization process. The goal was to digitize and preserve the original archival material for generations. This process included the selection of a digitization medium for the original archival material and of a widely used information system that would replace the existing propriety software for the metadata. Understanding the digitization process was vital. The creation of a reliable digital repository had to include several elements: the method for capturing the digital images (scanning or digital photography), the quality of the images (300 dpi or 600 dpi) and the conversion processes for resizing images, choosing file formats and renaming images. The resulting digital database would include quality images[13] created in a timely way and in a cost effective manner.

The methodology for capturing the digital images was selected: digital photography with 600 dpi of high-quality images.

The first question regarding the selection of an information system was whether this should be an automated system for archives or any other library system. In order to fulfil this condition, the selected system had to support the following features:

- the capability to support the data migration from a private Oracle-based system into a widely supported automated system;
- the possibility to export data in different formats such as: xml, Dublin core, MARC, allowing maximum flexibility and export to other systems, software, image viewers and discovery layer and exhibition programs;
- the functionality to preserve and enrich the links between records;
- the option to link the metadata to the digitized images;
- the capability of configuring Digital Right Management (DRM) for displaying and managing copyright messages and restricted images; and
- the integration of full-text search functionality of the archival documents.

The selected information system for this project was a library system which supported all the functionality mentioned above; the Aleph Integrated Library System (ILS) was selected for the migration and metadata management along with Aleph Digital Asset Module (ADAM) for the digital objects management including the option for DRM and full-text search[14]. An added advantage of Aleph was the fact that the ILS system has been used for several decades by the Hebrew University Libraries, is well known and supported by the existing staff – a fact that would considerably reduce the cost of ownership of the archive in the future.

3. Data management. Each specific item in the archive was catalogued separately – even if the content was identical. For example, an autograph letter and its typed version each had their own cataloguing record, which included a set of significant links between the two records. The new system enabled the user to retrieve all the manifestations of the document, to navigate between the records, creating a virtual folder of "associated documents" (see Figure 2).

An additional level of enrichment of the archive was the creation of another virtual folder which would include all related documents to a specific publication – such as the various reviews of an article, letters relating to the same publication and so on (see Figure 3).

The flexible configuration and infrastructure of the Aleph ILS allowed the migration of the archives complex configuration while preserving the metadata and allowing the creation of the virtual folders described above.

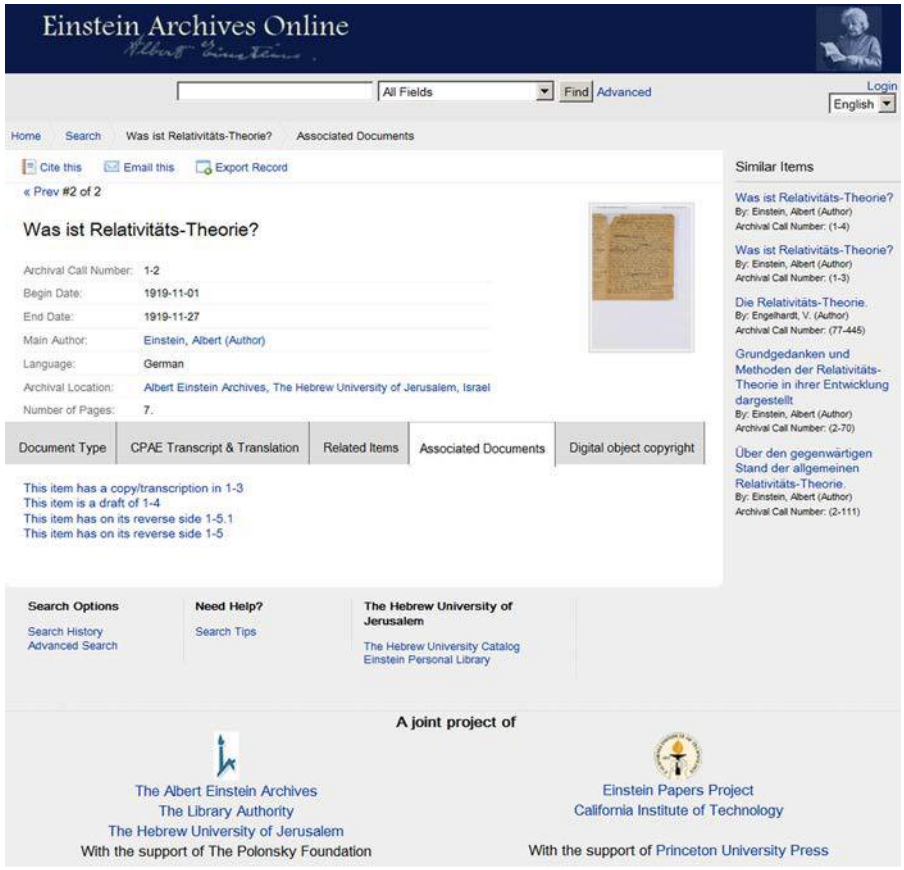


Figure 2.
The virtual archive – associated documents

4. Online access. The access to the original letters, manuscripts and documents was analysed within the framework of the technological advances for displaying digital collections. The goal was to achieve the display of the digital documents and also make these documents searchable. Tzoc (2012) explains how to select a document viewer for non-born-digital files based on the file format and the viewer properties. The selection of an image repository viewer was a must for this project; the viewer's features would include the display of images, the preview of individual pages of digitized multi-page objects, and the manipulation of images (such as zooming, rotating and turning pages).

The viewer would have to allow a fast and easy retrieval for all documents, even those which include hundreds of pages, such as notebooks and diaries. The viewer would also allow DRM functions for the restriction and copyright management.

The Open Library BookReader[15] was selected for this project. This open source software was also implemented in various other digitization projects such as the New York Philharmonic Digitization Archives[16] among others. The Open Library BookReader allowed the display of the digital images in an aesthetic way, the fast download of the high-quality items and the easy display to the end user with the ability of key interactions of the user as mentioned above.

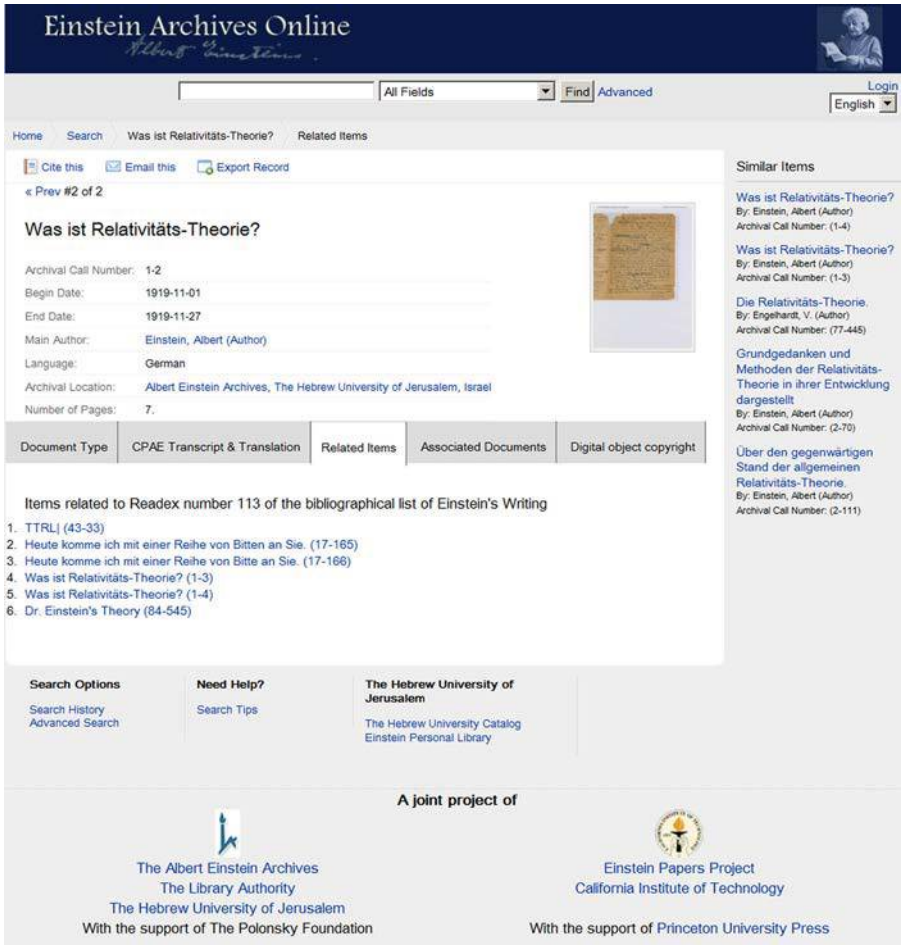


Figure 3.
Digitization and metadata processing

From the searchable point of view, the idea of applying optical character recognition to Einstein's original handwriting was a challenge; still, the goal of offering the possibility of full-text search from those documents was important. In order to achieve this goal, document transcripts and translations were loaded, indexed and linked to the original material. In this way, hundreds of items were made available for full-text search. The transcripts in German and the English translations were extracted from The CPAE[17].

The official Albert Einstein web site (http://alberteinstein.info) (see footnote 3) which was released in March 2012 included the possibility of free viewing and browsing of approximately 7,000 high-quality digitized images of Einstein's writings.

5. Navigation and discovery – the user experience. The goal of engaging all audiences is a top priority when planning a digitization project. The implementation of a friendly user interface to the archives would provide a good user experience taking into consideration all ages and level of expertise. It was important to adopt a discovery tool that would enable easy navigation into the archival material.

Vufind[18], an open source next-generation library catalogue interface, was selected for this purpose. The experience gained by implementing Vufind for the Hebrew University Libraries reinforced the idea of implementing this software for the Einstein Archives – enabling the use of the same discovery layer for the "One Archive and Eight Libraries" of the institution (Figure 4).

The implementation of Vufind for the archival material enabled the manipulation of the data exported from Aleph and the display of the digital images loaded in ADAM. Additional scripts were prepared in order to support the records links and metadata enrichment as well as the full-text search implementation in ADAM and the display of the digitized images using the Open Library BookReader viewer. All the strategic choices led to the achievement of the goals of the project plan including preservation, enrichment and accessibility to the original archival material.

The project launch

1. *Technical environment*. The search of the keyword "Einstein" in Google AdWords retrieves approximately about six million hits a month as seen in Figure 6. This was a good indication that the launch of the Einstein Online Archives would create a high amount of traffic.

Taking this information into account, the system presented by the Hebrew University to the world was strategically planned with a capacity to support daily millions of hits to the official web site and archival online database back on 19 March 2012, at the time of the launch of the project (Figure 5).

The technical aspects of the project launch included the task of creating an image of the local server (including the self-contained environment of the web site, archival database and digital repository) from the Hebrew University of Jerusalem. The image was packaged and transferred to an outsourced environment (Figure 6).

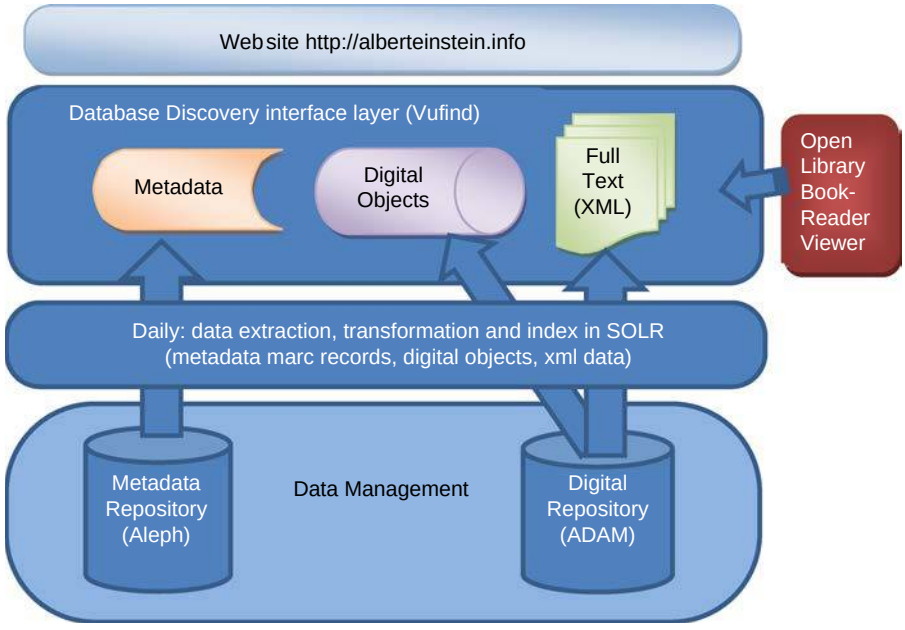


Figure 4.
Project implementation diagram

Figure 5.
Project gallery

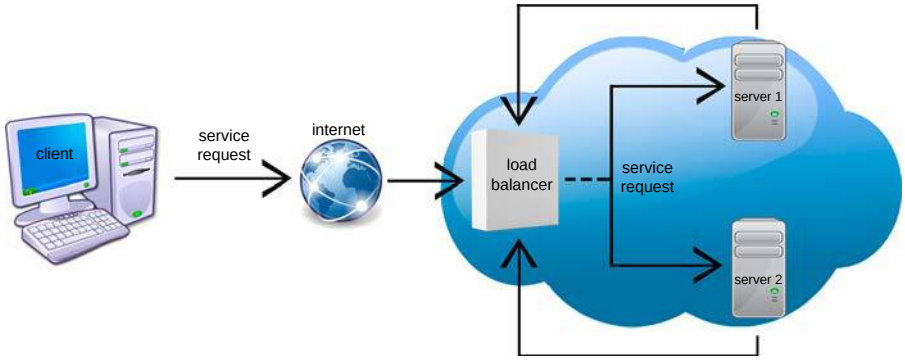


Figure 6.
Outsourced cloud environment

2. Performance. The Apache JMeter™[19] desktop application was selected for test performance in order to simulate the heavy load on a server and network. These tests offered the option of tuning the system for its strength, analysing the overall performance under different load types. This simulation was performed at the Hebrew University before the system was extracted for outsourcing. The system was also tuned in the outsourced environment. The stress test script was based on a configuration saved in an XML file. The scenario for testing was based on a specific search string. The test included 500 queries a second; this was done thousands of times with different search terms. The benchmarking was performed from Jerusalem, Israel and from California, USA.

3. Statistics. Monitoring the statistics[20] of the first five days after the launch revealed 21 million hits performed from 172 countries by over half a million unique visitors, an heterogenic audience with different interests, ages, learning approaches

with the one common denominator – an interest in the original work, diaries, notebooks, correspondence of Albert Einstein. This conclusion is directly drawn from analysing the search keywords and landing pages in the archival database. Popular searches were about relativity, physics and physics-related subjects such as "e = mc2", "light", "gravity", "quantum", "Brownian movement", "photoelectric effect", "1905", etc. Others were interested in the personal life of Einstein, the "love letters" he wrote, "Mileva" – Einstein's first wife, "Elsa" – his second wife, his "mother", "marriage" and more. Others, interested in the public life used search keywords such as "peace", "religion", "God", "socialism", "Jews", etc.

The most popular pages visited on the web site are as presented in Table I: the homepage, the archival database and the gallery. It was interesting to observe that the Finding Aid page – which includes Albert Einstein's timetable, the table of contents of the archive as well as a general description of the archive – did not get a high number of hits. Downloads of documents from the archival database are restricted by copyright and any request for publishing or commercial purposes must be addressed to the AEAs at Hebrew University.

The system performed in an excellent manner – no downtime, no interruptions or slowness. This successful outcome was due to the most positive collaboration between the Library Authority and the Computer Authority at the Hebrew University which produced a self-contained system (web site and database in a closed environment).

4. Online visibility. The Hebrew University of Jerusalem organized a press conference at the time of the project launch on 19 March 2012. The press coverage delivered over 420 articles and broadcast media indicating the access to 80,000 Einstein and Einstein-related documents including 2,000 full-text documents (7,000 images of the original archival collection). The project launch information displayed five consecutive days on Google Science News.

Table II provides a glimpse into the press coverage at the time.

General statistics 19 March 2012-9 May 2012	
Total number of hits	36,986,044
Requested pages: page requests ordered by hits = homepage	4,101,830
Requested pages: page requests ordered by hits = archival database	1,362,726
Requested pages: page requests ordered by hits = gallery	522,851
Number of unique visitors	594,227
Number of countries	172
Number of languages	129

Table I.
Statistics

TV	Online articles	Blogs
Al Arabyia	Jerusalem Post Haaretz	Mashable
ABC News	Time	Topsy
Fox News	*The Guardian* – UK	Physics News
	PC Magazine	Barnes&Noble
Channel 2 – Israel	Buenos Aires Herald	Compukiss
Sky News And more …	French Tribune And more …	EU Science News And more …

Table II.
Digitization project – press coverage, 19 March 2012

Conclusion: Einstein in the digital era

The digitization of the AEAs was very timely and fitted in very well with the new environment of "memory institutions", using the increasingly popular internet and Web 2.0 technologies. The great amount of "ordinary" people interested in the life and times of one of the best-known scientist of all times would now be able to access material previously accessible to privileged groups and scholars only. Thus, user expectations and information seeking behaviour had to be evaluated and taken into consideration when planning the digitization project as well as what had been the main consideration of digitization projects till then – the digital repository and its preservation.

One of the main lessons of the project was that digitization projects can be successful when a substantial budget is assigned along with the planning of human resources and staff expertise. These digitization projects require not only the archival staff expertise but also information specialists, IT staff, analysts and usually the digitization staff for processing the archival material. Finding the common language between all the professionals involved as well as building a good strategic plan are the keys to a successful project.

The AEAs digitization project, backed by the strong organization of Hebrew University of Jerusalem, the knowledge of the EPP in Caltech, California and the financial support of the Polonsky foundation was successfully launched and very well received. And one hopes that somewhere, among the many patrons of the archive, a spark of the imagination that according to Albert Einstein himself "is more important than knowledge […][21]" might be lit.

Notes

1. Prof Ben-Sasson is the president of the Hebrew University of Jerusalem. This quote is from the press conference at the launch of the digitization project on 19 March 2012.
2. An acronym designing the three types of institution – Library – Archive – Museum. Some – such as Wikipedia – go even further and talk about GLAM (G stands for galleries).
3. "HangingTogether is a place where some of the staff at OCLC Research, particularly those of us who support the OCLC Research Library Partnership, can talk about the intersections we see happening between these different types of institutions. We visit partners, go to conferences and take note of the interesting things we see along the way. Stop in, stay awhile, and hang out" http://hangingtogether.org/ (*Hangingtogether*, 2012)
4. The letter can be found at: http://ec.europa.eu/information_society/activities/digital_libraries/doc/letter_1/index_en.htm
5. The Europeana portal can the found at www.europeana.eu
6. "Derrida's attempt to deconstruct the concept of the archive – to explore the dynamics of the natural, psychological archive (the memorial archive) as well as the artificial, technological archive (the scriptural archive) – identifies a number of interesting characteristics of the archive or the process of archiving. Derrida's insights about the archive, which in later works he extended more fully into the digital age, can be synthesized constructively with more traditional or technical understandings of archives" (Paulus, 2011).
7. In the last decade, the universal adoption of the "big deals" with publishers resulted in almost identical collections of periodicals in most university libraries in the western world. The special collections which until then were mostly hidden in obscure corners became the focus of the library uniqueness.

8. ARL is a nonprofit organization of 125 research libraries at comprehensive, research-extensive institutions in the USA and Canada that share similar research missions, aspirations and achievements. www.arl.org/arl/index.shtml
9. The Polonsky Foundation is a UK-based foundation which financed, for example, the digitization of the Isaac Newton Papers, at Cambridge University. Dr Leonard Polonsky, from the foundation said: "This project unites the Hebrew University Library with digitization projects of the Polonsky Foundation recently launched at Oxford University's Bodleian Library and Cambridge University Library." This quote is from the press conference at the launch of the digitization project on 19 March 2012.
10. Archival material as cultural heritage: creating electronic versions and making it available online for work, study or research for future generations.
11. Archival material as science information: making research findings available online and keeping them overtime
12. Steven Buchanan discusses the different models for strategic planning and their importance applied to libraries projects (Buchanan, 2010).
13. Joint Information Systems Committee (JISC) at the UK published a tutorial for digitization projects (Quality Assurance Handbook: For JISC Digital Library Programs). This tutorial deals with digitization issues to take into consideration. The digitization tender at the Hebrew University included many of the elements mentioned in this tutorial.
14. Both systems are Ex-Libris products, http://exlibrisgroup.com
15. The Internet Archive BookReader – open source used for viewing digital images in libraries and archives (http://openlibrary.org/dev/docs/bookreader).
16. The Internet Archive Bookreader was implemented at the New York Philharmonic Archives database (http://archives.nyphil.org).
17. The Einstein Papers Project (www.einstein.caltech.edu) is engaged in one of the most ambitious scholarly publishing ventures undertaken in the history of science. The Collected Papers of Albert Einstein provides the first complete picture of Einstein's massive written legacy (Published by Princeton University Press – http://press.princeton.edu/einstein).
18. Open source for library catalogue developed by Villanova University and widely used by other academic libraries. More details can be found in: http://vufind.org
19. The Apache JMeter™ desktop application is open source software, designed to load test functional behaviour and measure performance (http://jmeter.apache.org/).
20. The statistics were created from the outsourced environment. The number of entries was calculated taken in consideration unique values, though the system performed with the load-balancing servers. The system was also monitored with Google Analytics which was not always accurate due to the high amount of traffic.
21. "Imagination is more important than knowledge. Knowledge is limited. Imagination encircles the world." Quoted in an interview by G.S. Viereck, 26 October 1929. Reprinted in "Glimpses of the Great" (1930).

References

Bolick, C.M. (2006), "Digital archives: democratizing the doing of history", *International Journal of Social Education*, Vol. 21 No. 1, pp. 122-134.

Buchanan, S. (2010), "Planning strategically, designing architecturally: a framework for digital library services", in Woodsworth, A. (Ed.), *Advances in Librarianship*, Vol. 32, Emerald Group Publishing Limited, pp. 159-180.

Chesbrough, H. (2003), *Open Innovation: The New Imperative for Creating and Profiting From Technology*, Harvard Business School Publishing, Boston, MA.

Chowdhury, G. (2010), "From digital libraries to digital preservation research: the importance of users and context", *Journal of Documentation*, Vol. 66 No. 2, pp. 207-223.

Dempsey, L. (2000), "Scientific, industrial, and cultural heritage. A shared approach: a research framework for digital libraries, museums and archives", *Ariadne*, available at: www.ariadne.ac.uk/issue22/dempsey (accessed 17 October 2012).

Derrida, J. (1995), *Archive Fever: A Freudian Impression*, University of Chicago Press, Chicago, IL.

Hangingtogether (2012), available at: http://hangingtogether.org/ (accessed 17 October 2012).

2005 July RLG Members Forum (2005) 'Libraries, archives, & museums – three-ring circus, one big show?", available at: http://web.archive.org/web/20060921021949/; www.rlg.org/en/page.php?Page_ID=20521 (accessed 17 October 2012).

Kirchhoff, T., Schweibenz, W. and Sieglerschmidt, J. (2008), "Archives, libraries, museums and the spell of ubiquitous knowledge", *Archival Science*, Vol. 8 No. 4, pp. 251-266.

Manoff, M. (2006), "Materiality of digital collections: theoretical and historical perspectives", *Portal: Libraries & The Academy*, Vol. 6 No. 3, pp. 311-325.

Marty, P. (2008), "An introduction to digital convergence: libraries, archives, and museums in the information age", *Archival Science*, Vol. 8 No. 4, pp. 247-250.

Nonaka, I. (1994), "A dynamic theory of organizational knowledge creation", *Organization Science*, Vol. 5 No. 1, pp. 14-37.

Oliver, A.L. (2004), "On the duality of competition and collaboration: network-based knowledge relations in the biotechnology industry", *Scandinavian Journal of Management*, Vol. 20 No. 1, pp. 151-171.

Oliver, A.L. (2009), *Networks for Learning and Knowledge-Creation in Biotechnology*, Cambridge University Press, Cambridge.

Panich, J.A. (2001), *Special Collections in ARL Libraries: Results of the 1998 Survey Sponsored by the ARL Research Collections Committee*, Association of Research Libraries, Washington, DC.

Paulus, M.J. (2011), "Reconceptualizing academic libraries and archives in the digital age", *Portal: Libraries & The Academy*, Vol. 11 No. 4, pp. 939-952.

Powell, W.W., White, D.R., Koput, K.W. and Owen-Smith, J. (2005), "Network dynamics and field evolution: the growth of interorganizational collaboration in the life sciences1", *American Journal of Sociology*, Vol. 110 No. 4, pp. 1132-1205.

Rayward, W.B. (1998), "Electronic information and the functional integration of libraries, museums and archives", in Higgs, E. (Ed.), *History and Electronic Artifacts*, Oxford University Press, Oxford, pp. 204-227.

Sutton, S. (2004), "Navigating the point of no return: organizational implications of digitization in special collections", *Portal: Libraries and the Academy*, Vol. 4 No. 2, pp. 233-243.

Tzoc, E. (2012), "Document viewers for non-born-digital files in DSpace", *Texas Digital Libraries*, Vol. 13 No. 1, p. 151, available at: http://journals.tdl.org/jodi/index.php/jodi/article/view/5600/5866 (accessed 3 January 2013).

West, J. and Gallagher, S. (2006), "Challenges of open innovation: the paradox of firm investment in open-source software", *R&D Management*, Vol. 36 No. 3, pp. 319-331.

Zorich, D.M., Waibel, G. and Erway, R. (2008), "Beyond the Silos of the LAMs: collaboration among libraries, archives and museums. Report Produced by OCLC", available at: www.oclc.org/research/publications/library/2008/2008-05.pdf (accessed 3 January 2013).

Further reading

Reilly, J.M. and Frey, F.S. (1996), *Recommendations for the Evaluation of Digital Images Produced From Photographic, Microphotographic, and Various Paper Formats*, Rochester Institute of Technology, Rochester, NY.

Zhou, Y. (2010), "Are your digital documents web friendly? Making scanned documents web accessible", *Information Technology & Libraries*, Vol. 29 No. 3, available at: www.ala.org/ala/mgrps/divs/lita/publications/ital/29/3/zhou.pdf (accessed 3 January 2013).

Corresponding author
Dalia Mendelsson can be contacted at: daliam@savion.huji.ac.il