

Digital Chronofiles of Life Experience

Cathal Gurrin^{1,2(✉)}, Håvard Johansen², Thomas Sødning³, and Dag Johansen²

¹ Insight Centre for Data Analytics, Dublin City University, Dublin, Ireland

cathal@gmail.com

² UIT The Arctic University of Norway, Tromsø, Norway

³ Oslo University College, Oslo, Norway

Abstract. Technology has brought us to the point where we are able to digitally sample life experience in rich multimedia detail, often referred to as lifelogging. In this paper we explore the potential of lifelogging for the digitisation and archiving of life experience into a longitudinal media archive for an individual. We motivate the historical archive potential for rich digital memories, enabling individuals' digital footprints to contribute to societal memories, and propose a data framework to gather and organise the lifetime of the subject.

1 Introduction

Earlier societies have left legacies informing us of their actions. Society has a need to record and document its events and has used whatever means possible. From Newgrange, Stonehenge and the Pyramids in Egypt, to the headstones on Easter Island, we are left with clues from a distant society. In many ways these societies created durable objects that tell a story that transcends time and sometimes these legacies leave us with a mystery to be solved in modern times. Historically the interaction and communication between individuals bore witness to events, and such material is today studied by historians and archivists. Whether stone carvings, exchange of letters, recorded phone calls, or the eyewitness details, this is what leaves the historical trace and forms the basis of the historical record.

As technology evolved, so too have the recording tools; hammer and stone, pen and paper, have been replaced by computers and digital recording devices. The ease at which we can create the historical trace is ever increasing and our ability to create an *evidence of me* has no bounds. Continual advancement in sensing technologies has lead us this point at which it becomes possible, should one wish, to continually record all of life activities into a personal media-rich archive, or as we refer to it in this paper, a *personal chronofile*. This process is called lifelogging and can quickly generate terabytes of information about the individual, in particular using wearable sensors and cameras, which are already flourishing on the consumer market [7].

Compare this to the most detailed life chronicle from the past, the Dymaxion Chronofile [18] where Richard Buckminster-Fuller documented his activities in detail day-by-day into a lifetime archive. Buckminster-Fuller referred to his archive as a “very accurate record of a human being”, consisting of 140,000

papers and 1,700 hours of audio and video. It is our conjecture that as lifelogging becomes a normative activity, archivists will have access to rich and unimaginably detailed records of many such individuals, not just those with the time or resources to manually curate a lifetime archive. These chronofiles would provide a first-hand, non-interpreted account of the past from people whose lifelogs provide a direct connection to historical activities and events.

In this paper we suggest that the chronofile generation process of lifelogging can revolutionise the concept of the historical archive. With minimal overhead, any individual will be able to maintain a chronofile. There are many challenges to be solved along the way, some of them documented in this paper, but the early adopters have shown that this is possible and as the benefits of lifelogging become apparent, it is likely to become the norm, rather than the exception. We propose a model of life-long personal chronofile management that takes the onus off manual curation, storage, and access to the digital objects by employing a set of transformation rules that filter or modify the digital objects.

2 Lifelogging and Life-Long Preservation

2.1 Personal Lifelogging

Lifelogging represents a phenomenon whereby individuals can digitally record their own daily lives in varying amounts of detail and for a variety of purposes. It offers great potential to mine or infer valuable knowledge about life activities. An example of a lifelog archive is shown in Figure 1, in which the activities of the lifelogger (a sequence of about 3,000 images) on the 29th May 2006 are organised automatically into a small set of logical events. Typically, early adopters to lifelogging considered it to be an activity for their own benefit. If lifelogging becomes more pervasive, one can imagine that many users would be willing to share aspects of their lifelog [17] with friends and family while alive or with archivists and researchers when passed away.

Lifelogging becomes possible as a result of three parallel advances in technology. Firstly, sensors are becoming cheap, reliable, robust, power-efficient and portable. There are many low-cost devices such as the OMG Autographer or the Narrative Clip that capture thousands of images per day. Digital audio recorders can record in detail 24/7 audio, a new generation of wearable video cameras can capture much of a day in high definition. Even our mobile phones include enough sensors to digitally lifelog our activities. It is now possible to gather a media-rich representation of our activities in detail that is previously unimaginable. Secondly, the cost of storing and transmitting large quantities of data from sensors and cameras has decreased. From [7] we know that we can currently store 6–8 years of wearable camera images on a \$100 hard drive. Thirdly, we have new search and artificial-intelligence techniques to allow us to convert large volumes of raw sensor data into meaningful semantic information that can derive new and aggregated knowledge from this data. All these three technological advancements have brought us to this point: the advent of the era of the lifelogs individual.

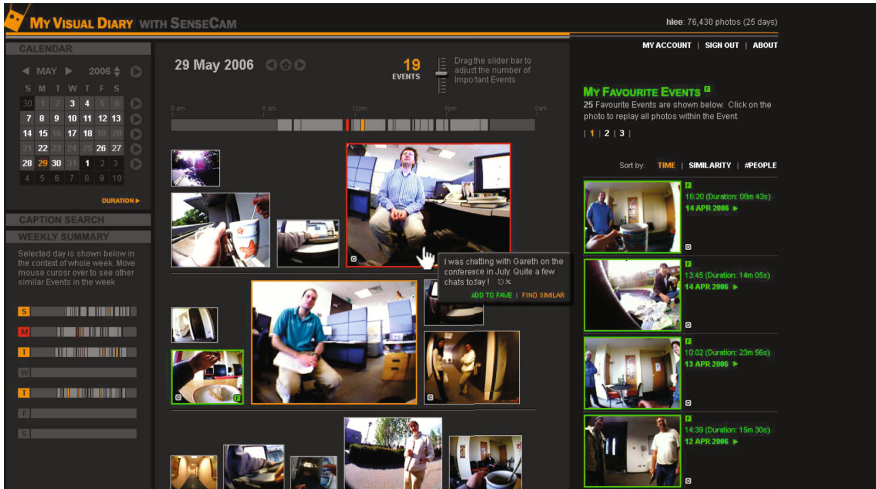


Fig. 1. An example lifelog browser [12]

There are many potentially life enriching benefits that could encourage an individual to engage in lifelogging, such as the potential for better self-awareness leading to longer and more active lifespans, new personalised healthcare applications, enhanced methods of learning, increased productivity in the workplace, increased independence, or mobility for people suffering from various memory and cognitive impairments, and new forms of offline and online social interaction [7]. It is considered that as these benefits become apparent, that lifelogging will become a normative activity.

Many use-cases of lifelogging have been proposed, for example, to assist the lifelogger in tasks such as personal healthcare, memory reminiscence [3], and browsing a digital record of past activities [12]. There is now real potential that we are on the cusp of an era of, what Gordon Bell and Jim Gemmell refer to as *total capture* [2]. In fact, this has already begun with the market availability of lifelogging devices.

2.2 Lifelogs as Societal Records

We know from [14, p.139] that “all events have their witnesses, their memory: the trace” and that witness is born through communication of events. The tools we employ for communication today far outweigh the pen and paper of yesterday. Chronofiles promise a new dimension to witness, memory and trace. Consider if you will, a historian in 2114; he will have access to vast chronofile archives from the deceased first generation of lifeloggers. He will have access to orders of magnitude more information about the individual than we are currently gathering in 2014. The events, as before, will be witnessed and memorised, but the form they leave in the trace is so much richer in a world of lifelogging. Mis-interpretation of events and generating historical narrative based on only snapshots of historical communications will be a problem of the past.

One example is the story [15] of the sinking of the “Olong” and the subsequent misinterpretation of events and witness accounts. Human error and distant reporting of a chaotic scene taking place far from land where a sinking boat and a father gesturing the Navy to take his child becomes the story of desperate refugees willing to use all means possible to force the Navy to rescue them, including risking the lives of their children by throwing them overboard. This event has been documented and analysed in detail. Witness statements and logs are placed within a kind of system-centric approach to documenting, which means that the story is documented from societies point of view, placed in system so that we can learn more about society. The story is about a group of people trying to make it to Australia for a better life, but we are missing the story about the individual. The individual’s search for a better life, the journey, the sinking and what subsequently happened to the individuals is lost. The true story lies distributed in peoples memory and portions of this story die as memories fade. The story of the individual takes second place to society’s story. In the era of chronofiles the story about the individual does not have to be a footnote in a formal document.

2.3 Data Formats

While most historians will find lifelogging a treasure trove of information, it will also certainly pose problems in terms of long term preservation that may not be obvious yet. When it comes to electronic material it is already clear that archivists have a lot of trouble maintaining collections of electronic material. The reason for this is the short lived timespan that some file formats have and their evolutionary nature. Today we even have trouble correctly interpreting some word-processing documents from twenty years ago.

3 A Life Digitised

To understand the potential sources of knowledge and evidence contained in chronofiles, we can look to the early adopters and the data that they store. The manually generated chronofile of Richard Buckminster-Fuller, that we mentioned earlier, is a detailed first-hand record of a life lived. It is unlikely that there are many others willing to put in so much effort. Recent early adaptors, such as Bell [1] and Mann [13] show the potential of what can be captured for a basic first-generation chronofile and they motivate the necessity for chronofiles to be generated automatically by wearing a small set of non-intrusive sensing devices. These first generation chronofiles, which are called lifelogs, are simple data capture, storage, processing, and interaction frameworks that store the data in fixed formats and do not take into consideration the archival challenges of long-term storage and curation of the data. Figure 2 shows an example architecture of a first generation, widely deployable lifelog that is concerned with capturing rich data about the user experiences; organising and enriching it to make it usable for the individual; and finally supporting pervasive access via current access devices.

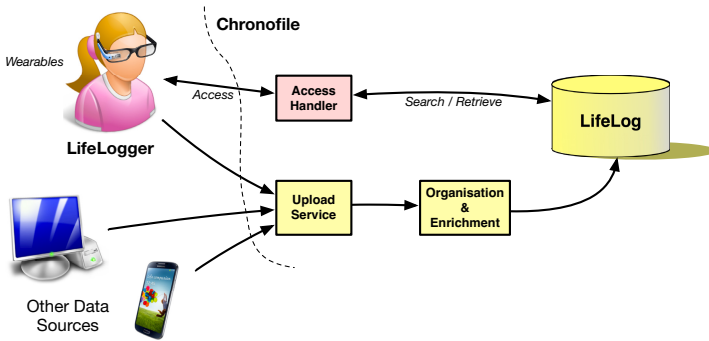


Fig. 2. The Structure of a Basic Lifelog Archive

3.1 Data Storage Requirements

The lifelogs being captured today already provide detailed knowledge about the minutiae of life in 2014, with many potential sources of information that can be included. The list of sources from [7] includes passive visual capture devices that take up to 4,000 images per day or short-durations of video, passive capture audio to record the sounds of life, personal biometrics to understand the physical state of the individual, mobile device context to understand the individuals place in the environment, the communications of the individual, the content they create and consume, their environmental context, media consumption, and any annotations to that content. To illustrate the variety of data sizes and quantities, a summary table of a selection of lifelog data is shown in Table 1. In this table we include the annual storage requirements as well as a lifetime (85 years) storage requirement.¹

All this data is inherently unmanageable if left to the individual. The huge data volumes involved require the development of a suite of organisation and search tools that allow the individual to gain benefit from the lifelogging process. At present, these search tools, described in more detail in [7] allow for the conversion of a raw chronofiles into a manageable set of digital objects. To take a human memory analogy, these roughly align with the idea of an event in the life of a human (see the event list in Figure 1). The digital objects then act as the unit of observation and retrieval for the chronofile. In order to be located by a search mechanism, these events need to be enriched with appropriate meta-data, indexed by search tools and made available via an appropriate interaction mechanism (e.g., Google Glass for the lifelogger or a desktop archive search tool for an archivist).

With the exception of digital video data, one can easily imagine storing a lifelog of an individual indefinitely. Because at current capture rates, it requires about 1 TB per year to capture wearable camera photos from an off-the-shelf

¹ Extrapolating across a lifetime, when there is an assumption that data bit rates remain static is merely for illustrative purposes only. Data sources, qualities, resolutions, and bit rates are constantly increasing.

Table 1. An illustration of the data quantities and data sizes for a selection of chronofile data in 2014, over a day, year and a lifetime (using 85 year lifespan and 16 hour days)

Content Type	Data/year	One year	Lifetime
4K Video	5,840 hours	342 TB	28.3 PB
HD Video	5,840 hours	32.8 TB	2.65 PB
Autographer Camera	1.1 million images	479.6 GB	40.8 TB
Audio (mono - 22 KHz)	5,840 hours audio	227.8 GB	19.4 TB
Accelerometer (1 Hz)	21 million readings at 1 Hz	0.05 GB	4.25 GB
Locations (0.2 Hz)	3.9 million GPS points	0.01 GB	1 TB
Reading Log	User dependent	1 GB+	80 GB

5 Mega Pixel wearable camera, along with various other continually sampling sensor data (e.g., accelerometer, GPS, and audio snippets). Consider the Autographer wearable camera. If one wore it for an 85 year lifespan, the storage requirements of six photos per minute would only require ten of today's hard drives. Today, it is therefore feasible to store this data permanently and provide direct non-interpreted knowledge about individuals.

Although the data produced by lifeloggers today is manageable with current technologies, information technology seemingly has a tendency to follow an exponential pathway of advancement. We have already seen this in Moore's Law for CPU transistor densities, Kryder's Law [4] for disk storage, and in Kurzweil's Law of Accelerating Returns [11] for general information technology advancement. Assuming lifelogging archive sizes follow similar growth curves, a capture rate of one Terabyte per year today has the potential to become 1 Zettabyte of data per year by 2045, or 1 Geopbyte by 2075.²

Some might argue that ever growing storage device capacities coupled with efficient compression algorithms will adequately support any future storage demands of personal lifeloggers. Others argue the contrary; that the modern society already produces way more data than available storage can handle. In particular, IDC [5] already predicted in 2007 that the 255 Exabytes of information created and replicated that year was for the first time to surpass the 246 Exabytes aggregate storage capacity available globally. Also, they predicted that the gap between disk demand and supply would just be larger. Some years later, their prediction may be correct.

3.2 Curation

Expecting an individual to curate their chronofile in a fine-grained manner to fit within available storage space, to migrate to new file formats, to address potential legal and privacy issues, is not realistic. Even today, the early adopters see chronofile archives as a form of data dump into which all possible data is

² This only assumes a simple lifelog configuration including the equivalent of an OMG Autographer and various other forms of sensed media. Were one to consider recording 4K video 24/7, then the storage requirements increase by an order of magnitude.

stored, either because it is too troublesome to manually curate the data, or because it is considered more prudent to keep all data on the chance that it could potentially be useful at some point in the future. Therein lies the problem that we foresee; the chronofile needs to be self-organising. We therefore propose a model for chronofiles, based on the Gardi model [6] we proposed for multimedia collections, which iterates over data sources temporally and applies higher-order information transformation rules directly on the stored digital objects. Gardi was positively evaluated for security video archives, home video, recordings and a very early stage visual lifelog.

In previous work, we have successfully applied such application specific data curation techniques for a concrete lifelog scenario in the sports domain. A lightweight lifelogging service was built and deployed for head coaches as part of elite soccer performance development. Soccer arenas and training grounds are populated with stationary [9] and portable cameras [8], and mobile devices provide support for hindsight annotations of sequences worth capturing. This also includes prototypes with Google Glasses used on the field by coaches. This way, coaches determine after a certain event has unfolded whether to permanently store this or not, and, if so, touches the cellular screen. Experience from a complete soccer series in Norway for a specific team over a year indicates that about 5% of video footage from complete games are persisted this way and made use of for analytics and feedback procedures. The other 95% of the video footage rendered useless with regard to events worth examining, can still be stored somewhere for archival purposes or permanently deleted. We conjecture similar techniques have broad applicability for personal chronofiles and long-term preservation.

3.3 Privacy

Since we do not envision that chronofiles will be curated by individual users, we suggest that policies regarding use and curation will be of concern. This implies that chronofiles also have the potential to be damaging if accessed inappropriately during the individual's life, or if accessed insensitively after death. Consider the case of the Norwegian explorer Fridtjof Nansen; he was an explorer, a scientist, a diplomat, a humanitarian and Nobel Peace Prize laureate. In 2011 a book was published detailing explicit exchanges between Nansen and his much younger partner. His once private collection had been put on display. This resulted in a change in the perception of Nansen, and is testament to the fact that the story the archive tells about a person can be damaging. A chronofile, assessed after an individual's death may contain deeply personal data.

Hence the need for automatic curation to ensure privacy and reputation would be very necessary. Privacy regulations and personal privacy constraints must be properly supported by the software hosting and maintaining these chronofiles. This must be provided for the computer-illiterate as a rule, not exception. Also, the chronofiles will be stored and maintained in a virtual network, not a centralised physical digital library, including enterprise silos, public services, specialised cloud providers, and even private utilities.

4 Chronofiles Transformation Model

Based on the ideas and discussions listed above, we propose an approach for lifelong and post-life management of personal chronofile data. Our model extends the basic lifelogging architecture described in Section 3 with the the concept of lifecycles of lifelog data using a set of transformation rules that operate over the chronofile data as needed, as illustrated in Figure 3. These transformation rules, in their present form, are classified as:

- *Robustness Transformations.* Maintain robustness of the data in the chronofile to hardware errors.
- *Migration Transformations.* Maintain data in currently accessible formats through periodic automated digital format migrations.
- *Enrichment Transformations.* Maintain modern and accurate metadata by means of new generations of semantic enrichment tools.
- *Storage Transformations.* Maintain an appropriate quantity of content by deleting content only if necessary, but keeping as much rich detail as possible, for example, by deleting repetition, but keeping novel content.³
- *Legal Transformations.* Protect the individual owner, and people captured in the archive, by adhering to data storage and privacy laws⁴ as they apply to the individual.
- *Reputation Transformations.* Protect the reputation of the individual after control of the chronofile is relinquished, for example, after death.

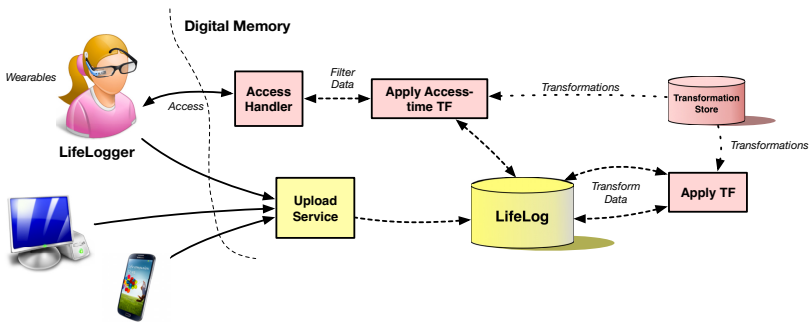


Fig. 3. An initial model of a Chronofile (TF = Transformation)

These data transformations can be independent tools that are automatically executed as required by the chronofile management software and permanently modify the data. There would also be access-time transformations, the aim of

³ For more information on transformation rules for lifelogs, see the initial set in [6].

⁴ Privacy and the Right to Privacy are concepts that vary across time and jurisdiction. In many cases, it may be more appropriate that the chronofile applies access-time, rather than permanent transformations on the data.

which is to limit what a user can see, while maintaining the integrity of the underlying data. For example privacy restrictions on lifelog data may require the chronofile to hide private or intimate content from some or all viewers, or display transformations which format chronofile data to take advantage of whatever access mechanism is being employed.

We have embarked on building a series of systems targeting long-term virtual storage infrastructure for early-stage personal chronofiles [6, 10, 16]. Our fundamental approach is to add self-contained and autonomic meta-code to the data collections or individual data items complementing traditional meta-data of today. In this way, expressive transformation or curation rules live alongside the data to be curated and can be maintained over a long timespans.

5 Conclusion

A traditional and important source of information when archivists and historians try to piece together the story of someones life is the subjects own diary. Combined with the contents of exchanges of information (letters) and other sources, the historian will carefully analyse and create an understanding of a person. In the traditional archive sense, there are few people that have their entire lives documented. The chronofile changes everything—forever. In many ways the chronofile reduces the influence of historians on the historical narrative. Historians will no longer have to fill gaps by matching various information pieces together, guessing and interpreting. This has the potential to give the individual a lot more control of how history will view them, but also the potential to have personal data examined. On the other hand the sheer volume of material requires the development of new tools and opens for many new exciting research areas for historians. Both inter and intra chronofile analysis will provide historians with an ability to understand the path-of-life in a manner that we have never been able to before.

We have discussed the data lifelogging that can generate and motivated why this could be a very useful data source for historians and archivists. We considered how lifelogs operate today and look forward to a time when the data quantity and richness of a lifelog will be orders of magnitude greater than today. We then proposed a model for chronofiles that allows them to grow and be managed by software, yet still maintain detailed representations of life activities and migrate automatically to latest formats, so that they are available for the lifelogger, and later the historians and archivists. Chronofiles gives the individual the ability to decide in advance what material should be available and to whom. Perhaps chronofiles will allow historians of the future to trace certain life-events of the individual and better understand their consequences for society as a whole. The analysis of multiple chronofiles within in social sciences setting may lead to a much better and deeper understanding of human behaviour. Chronofiles have the potential to positively impact on the life of the individual, the historian and society as a whole.

References

1. Bell, G., Gemmell, J.: A digital life. *Scientific American* **296**, 58–65 (2007)
2. Bell, G., Gemmell, J.: *Total Recall: How the E-Memory Revolution Will Change Everything*. Penguin Books (2009)
3. Berry, E., Kapur, N., Williams, L., Hodges, S., Watson, P., Smyth, G., Srinivasan, J., Smith, R., Wilson, B., Wood, K.: The use of a wearable camera, sensecam, as a pictorial diary to improve autobiographical memory in a patient with limbic encephalitis. *Neuropsychological Rehabilitation* **17**(4), 582–601 (2007)
4. Chip, W.: Kryder's law. *Scientific American* pp. 7–25 (2005)
5. Gantz, J.F., Reinsel, D., Chute, C., Schlichting, W., McArthur, J., Minton, S., Xheneti, I., Toncheva, A., Manfrediz, A.: *The Expanding Digital Universe*. White paper, IDC Corporate USA (2007)
6. Gurrin, C., Aarflot, T., Sav, S., Johansen, D.: Addressing the challenge of managing large-scale digital multimedia libraries. *Journal of Digital Information Management* **7**(5), 262–270 (2009)
7. Gurrin, C., Smeaton, A.F., Doherty, A.R.: Lifelogging: Personal big data. *Foundations and Trends in Information Retrieval* **8**(1), 1–125 (2014)
8. Johansen, D., Stenhaug, M., Hansen, R., Christensen, A., Hogmo, P.M.: Muiithu: Smaller footprint, potentially larger imprint. In: *2012 Seventh International Conference on Digital Information Management (ICDIM)*, pp. 205–214 (August 2012)
9. Johansen, D., Halvorsen, P., Johansen, H., Riiser, H., Gurrin, C., Olstad, B., Griwodz, C., Kvalnes, Å., Hurley, J., Kupka, T.: Search-based composition, streaming and playback of video archive content. *Multimedia Tools and Applications* **61**(2), 419–445 (2012)
10. Johansen, H.D., Zhang, W., Hurley, J., Johansen, D.: Management of body-sensor data in sports analytic with operative consent. In: *Proceedings of the 2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*. IEEE (April 2014)
11. Kurzweil, R.: The law of accelerating returns. In: Teuscher, C. (ed.) *Alan Turing: Life and Legacy of a Great Thinker*, pp. 381–416. Springer, Berlin Heidelberg (2004)
12. Lee, H., Smeaton, A.F., O'Connor, N.E., Jones, G.J., Blighe, M., Byrne, D., Byrne, D., Doherty, A., Gurrin, C.: Constructing a SenseCam Visual Diary as a Media Process. *Multimedia Systems Journal, Special Issue on Canonical Processes of Media Production* **14**(6), 341–349 (2008)
13. Mann, S.: Continuous lifelong capture of personal experience with EyeTap. In: *Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences, CARPE 2004*, pp. 1–21. ACM, New York (2004)
14. Matsuda, M.K.: *The memory of the modern*. Oxford University Press (1996)
15. McKemmish, S., Piggott, M., REED, B., Upward, F. (eds.): *Traces: Document, record, archive, archives by Sue McKemmish*. Centre for Information Studies - Charles Sturt University (2005)
16. Nordal, A., Kvalnes, Å., Hurley, J., Johansen, D.: Balava: Federating private and public clouds. In: *2011 IEEE World Congress on Services (SERVICES)*, pp. 569–577 (July 2011)
17. O'Hara, K., Tuffield, M., Shadbolt, N.: Lifelogging: Privacy and empowerment with memories for life. *Identity in the Information Society* **1**(1), 155–172 (2009)
18. Quimby, S., Mandeville-Gamble, S., Pang, A., North, A., Chiu, P., O'Hanlon, B. (eds.): *Dymaxion Chronofile*. No. M1090 in Collections, Stanford University (2006)