

Article

Labours of Love and Convenience: Dealing with Community-Supported Knowledge in Museums

Stefano Cossu

Software Architect, J. Paul Getty Trust, 1200 Getty Center Drive, Los Angeles, CA 90049, USA; scossu@getty.edu

Received: 3 January 2019; Accepted: 27 February 2019; Published: 7 March 2019



Abstract: This writing utilizes the case study of a specific project, namely adopting a Digital Asset Management System (DAMS) based on open source technologies at the Art Institute of Chicago (AIC), to describe the thought process, which along the way led to the discovery of Linked Data and more general technology development practices based on community participation. In order to better replicate such a thought process and its evolution into a broader strategy that goes beyond technology, this paper will begin by describing the problem that the Collection IT team at AIC had been initially tasked to resolve, and its technical implementation. After that, the paper will treat the strategic shift of resources from a self-contained production and review cycle toward an exchange-based economy. The challenges, both external and internal, posed by this change will be addressed. All the while, the paper will highlight perspectives and challenges related to the museum sector, and the efforts of AIC to adopt views and methodologies that have traditionally been associated with the library world. A section is dedicated to ongoing efforts of the same nature among museums.

Keywords: museums; community-driven knowledge; sustainable development; Linked Data

1. Introduction

In 2013, shortly after joining the Art Institute of Chicago (AIC)¹, I was tasked with implementing a Digital Asset Management System (DAMS)² for the AIC collections. At that time, the collection information systems at AIC consisted of one Collection Management System (CMS) that had been developed in house since 1991, called CITI. This system had been built originally in HyperCard, and then in the 2000s, migrated to 4D, a proprietary framework mostly known in the EU, which consists of a relational database, a programming language, an Integrated Development Environment (IDE), and a User Interface (UI) toolkit.

The reason for this somewhat unusual setup among museums, which today rely for the vast majority on more or less established CMS vendors, may in part lay in the AIC size and complexity of workflows, and in the fact that AIC invested early in the digitization of its cataloging tools and practices. Back in the early 1990s, there were very few CMSs, if ever, capable of managing the complexity of operations around the AIC's hundreds of thousands of accessioned items. Hence the need for a home-grown system, which has so far served its purpose well in terms of responsiveness to the museum staff's needs.

¹ At the time of this writing, the author is no longer working at AIC and is unaware of further developments of the project described, which was still evolving at the time of his departure. Therefore, all narration is in the past tense, even when it may describe situations that may currently persist.

² The acronyms DAM and DAMS are some times used interchangeably in other literature, often the latter as the plural form of the former. In this paper the term DAM is strictly referring to the abstract discipline of Digital Asset Management, while DAMS indicates a concrete digital system or systems for digital asset management.

CITI, overall enjoying high user approval ratings, had a few structural downsides. While AIC had two dedicated full-time developers to maintain the system, the project had a very long backlog of feature requests and bug reports. Several low-level features such as “virtual” tables and fields (corresponding to “Views” provided by most relational database systems) and an access control framework had been built ad-hoc early on, creating very deeply nested domain-specific features that are normally provided out of the box by most modern database and application development environments. Additionally, the very sparse adoption of 4D in North America, constantly shrinking due to the growth of open source programming languages and database systems, made finding developers for CITI a very hard task. Even when developers were found, the specific implementation of CITI presented a very steep learning curve even for experienced 4D engineers, to the point that it would take a very long time for them to gain a satisfactory confidence with the code.

The above factors, while presenting serious concerns, did not prevent CITI from being an irreplaceable day-to-day application for a very long period of time, especially considering the prospect of the enormous migration task that a replacement would entail. When the need for a DAMS took shape, however, it was clear that adding to the already backlogged CITI road map would not be sustainable. CITI had some basic image and interpretive media management tools built in, however developing a full-fledged DAMS within it would have been unthinkable. An external system had to be developed or adopted that would run alongside, and communicate with, the existing CMS. And that is when the largest of the legacy system’s limitations came to the foreground.

CITI had been, up to this point, a self-contained system. The advent of a major system demanding large amounts of complex data structures with very frequent updates, and exposing data back to CITI in order to provide the users with a smooth transition between collection record and digital asset management operations, combined with the need for identity management of users that were not managed by a central directory service, posed serious interoperability challenges. We knew that many factors in the existing architecture would be extremely expensive or impossible to change, so we needed to adopt a system flexible enough to work around such restrictions.

When the DAMS project started, a three-year timeline and a budget had already been allocated. This allocation was made under the assumption of the purchase of, and (smooth) transition to, a “turnkey” system rather than heavy custom development. Thus, after we realized that no *deus ex machina* “solution” could meaningfully resolve our problem, we decided that we had to work in a very efficient and frugal way to achieve our goals, while maintaining long-term sustainability goals.

2. Materials and Methods

2.1. Black Boxes and Hack Boxes

Aside from the user-driven requirements, the key internal and technical factors for choosing our DAM product were interoperability and flexibility. All commercial DAM solutions reviewed were lacking in either or both, be it for poorly designed (and even more poorly documented) APIs, or for the fact that we anticipated it to be very difficult and expensive to customize the default behavior of a pre-packaged DAMS to work with our particular situation. Additionally, our users were accustomed to CITI being able to accommodate most of their needs, since it was built in house. A vendor-provided product could have been perceived as too rigid for the users and would have been harder to accept.

At the same time, we did not want to continue reinventing the wheel by building a repository system from scratch. Even if we could afford to implement all the complex functions of storing, backing up, linking, and serving large amounts of files, along with their metadata, that would have been pointless with all the existing software that already resolved that same problem.

At that time we were only considering two options to our problem: “black boxes”, i.e., proprietary or open source technology, which is adopted or purchased with the assumption that it would “just work” and requires no in-house expertise, or “hack boxes”, i.e., technology that one has full control over, and whose maintenance burden would be completely on us.

There are good uses for both black boxes and hack boxes. Generally, one would want to write and maintain as little custom code as necessary, and only for institution-specific situations or ones that there is no good enough solution for on the market. While we used projects such as Apache, Python, Postgres, and Solr, which are fully open to community contribution of all sorts, any kind of contribution to those projects was far beyond our capabilities. Practically, at that point, “open source” for us simply translated into “free of charge”.

This is when we ran into Fedora³. Fedora was a good fit for our case in many ways: it provided the fundamental, hard to implement repository services, leaving complete freedom over the choice of front ends or integration tools, which is where we needed the widest latitude of decision. Additionally, the whole interaction paradigm with Fedora is designed around the API, not the other way around (some times as an afterthought) as we saw in other products.

After a period of study and evaluation, and numerous phone interviews with institutions who were running it (and who very generously shared information about it), we found Fedora to be a middle way between black boxes and hack boxes: It is open source, and fully community-supported—unlike some other Open Source Software (OSS) that is mainly supported by commercial entities and is exposed to the whims of the market. Moreover, Fedora is sustained by a much more specific community, very close to AIC, compared to the above mentioned general-purpose projects. It is built by and for academic institutions with a scholarly purpose and mindset.

Our production architecture, named LAKE, ended up being a mix of black boxes, hack boxes, and middle-ground solutions:

- Repository services: Fedora (community project)
- DAM: Samvera⁴ (heavily customized community project)
- ETL: Combine (hack box)⁵
- Indices: Blazegraph⁶, Solr⁷ (black boxes)
- Messaging and integration: Apache Camel⁸, ActiveMQ⁹ (black boxes)
- Web publishing: Marmotta¹⁰ (black box) with custom translation layer (hack box)
- Image server: initially Loris¹¹ (community & hack box), then Cantaloupe¹² (black box)

The pattern that emerges from this scenario is that hack boxes were adopted mostly as “glue” between closed systems, or where very AIC-specific issues needed to be resolved. Our process was identifying projects that would fulfill specific, isolated tasks, with good integration capabilities, and that ideally had comparable alternatives (in case they were to be replaced further on, which happened), and connect them using the least amount of custom code possible. The only large coding efforts were Combine, the Python ETL framework, which was necessarily tailored to the AIC Collections data model, and LAKeshore, the staff-facing DAM application, which was customized to our users’ workflows.

³ Fedora (<http://fedorarepository.org/>) is a Linked Data Platform repository service to store binary files and metadata.

⁴ Samvera (<http://samvera.org/>), formerly known as Hydra, is a Ruby On Rails application that provides UX and workflow management on top of a Fedora repository.

⁵ Combine is a custom Extract, Transform, Load (ETL) application written in Python and tailored to the AIC data model.

⁶ Blazegraph (<https://www.blazegraph.com/>) is a triplestore, i.e., a database for RDF, which is the language to formally represent Linked Data. The project is discontinued as of 09-02-2019. A similar, cloud-based product called Neptune (<https://aws.amazon.com/neptune/>) is owned by Amazon.

⁷ Solr (<http://lucene.apache.org/solr/>) is a high-performance document store optimized for full-text searching on unstructured or lightly structured data.

⁸ Apache Camel (<http://camel.apache.org/>) is an integration framework, that watches events (e.g., creating or updating resources) that Fedora stores in ActiveMQ (see next note) and uses these events to trigger actions (e.g., updating an external index).

⁹ ActiveMQ (<http://activemq.apache.org/>) is a message queue manager.

¹⁰ Marmotta (<http://marmotta.apache.org/>) provides a pipeline to query and transform data from an LDP server (Fedora) using a “transform program” language called LDPPath.

¹¹ Loris (<https://github.com/loris-imageserver/loris>) is a IIIF-compatible image server written in Python.

¹² Cantaloupe (<https://medusa-project.github.io/cantaloupe/>) is a Java-based IIIF image server.

The modular nature of this architecture, while complex to set up and initially to handle, proved its advantage when issues with the image server arose. Initially, we had invested heavy development efforts in Loris, a Python-based IIIF server. The team was proficient in Python, therefore the adoption of Loris seemed like the best choice, since we needed to customize the way images were retrieved from Fedora and served to the public. However, when stability issues arose with the production server, we decided to replace it with a system that we did not need to customize: We introduced Cantaloupe, a server that operates using the exact same protocol (IIIF) as Loris. Cantaloupe is written in Java, but we were only interested in customizing the image resolver, not the whole software. The resolver used Ruby scripts, which were easily crafted. Within a week, which was mostly spent testing, the image server was replaced in production.

An opposite example (going from a black box to a hack box) is when we realized that one other component, Marmotta, had stability issues. Marmotta had not been updated in several years, so requesting feedback was hopeless. Our engineer developed then a custom server in Python and ANTLR¹³ within a few days which was meant to replace Marmotta. The replacement had not yet happened when I left my position at AIC, however that was simply due to prioritization; reportedly, the software did exactly what was needed, not more, not less.

Implementing LAKE was a very deep and informative experience for the whole team, not without its challenges and tensions. Adopting a model that had no precedent in museums and very different approaches in other cultural heritage institutions was a great challenge. Four years after the project start, the architecture was running in production with the core image set (about 350,000 images and text documents) migrated from CITI and various file archives, and more assets slated for migration in subsequent phases. The system synchronized CITI, LAKE and the public-facing mirror every 10 min, for up to tens of thousands of record updates a day. Its triple-store index held over 124 million data points. The biggest issue with the system were performance bottlenecks, mostly related to CITI-LAKE syncs and publication of resources; on the other hand, almost all of the components (except for the above mentioned Marmotta) were exceptionally stable once configured correctly, practically without crashes.

We paid a price for implementing Linked Data at the lowest level of LAKE, especially when the scope of the system shifted. Initially the plan was to have LAKE as the main aggregation point for both internal and published data, and to expose a SPARQL endpoint as well as a Solr index to the Web; however, with a change of department layout, other tools were developed to index and aggregate published data, the scope of LAKE changed, and we eventually found ourselves doing a lot of data movement and transformation for no good reason. We then decided that we would change the sync process so that CITI would be directly responsible for pushing its data to the public endpoint, while LAKE would only retain a minimum set of CITI data necessary to establish connections between assets and CITI entities, while it would only publish asset-related data.

The positive note about these major changes is that they are very unlikely to be disruptive, given the modular structure of LAKE. While the complexity of the system required a significant effort to set up, the system was able to respond to significant shifts in the architecture and even to the replacement of whole components.

2.2. A Cultural Shift

By adopting Fedora, we discovered a much more neighborhood-like community than what we had been used to: In this community each participating institution and individual has their own weight. While Fedora is written in Java, for which we had no in-house expertise nor interest to invest efforts in, AIC contributed to Fedora by providing use cases, participating in technical and strategic discussion, and testing candidate releases. As the only major museum in the Fedora community, AIC's

¹³ <https://www.antlr.org/> (accessed on 02-15-2019).

angle on topics that so far had been mostly viewed from a librarian's perspective was particularly important to expand the capabilities of the software and the vision of the project. On the other hand, our team's exposure to topics that libraries have had a firm grasp on for a long time, greatly enhanced our understanding of digital repositories beyond the common museum DAM use cases.

With our involvement in the Fedora community came the realization that our team's focus and mode of operation needed to change. Our project life cycle had so far been managed through a small team of developers, occasionally aided by external contractors, who had a full decision power over the technical implementation. With community-supported software, any improvement or bug fix needed to be coordinated with a larger cross-institutional group, with each participant responding in part to the Fedora community, in part to his or her own institution. My responsibility as a manager, formerly limited to coordinating a contained number of engineers based on institutional stakeholders' requirements, now demanded interaction with groups outside AIC, who made tactical and strategic decisions collectively.

The reward for this added complexity was that the Fedora technical and governance teams achieved much more than our team would have been able to by itself. Additionally, the exposure to common problems and their collective resolution greatly enhanced our team's knowledge of the matter.

With our contribution came the opportunity to participate in strategic decisions. Having a seat in the Fedora Leaders and Steering groups allowed AIC to have a voice in shaping the direction of Fedora. This was an aspect close to the hack box scenario, without the full maintenance and responsibility burdens. Furthermore, the fact that we did not have full decision power, i.e., that our propositions were discussed with a group, was a positive factor because, many times, our ideas were augmented, improved, or redirected to better solutions, rather than being implemented in a vacuum.

This kind of synergy only happens in a community whose goals are very close to one's institution's. This was the case with Fedora, where discrepancies between stakeholders were relatively small, which could also be because Fedora is quite neutral, and in its own sense, a general-purpose tool. We had a somewhat different experience with other projects, for example Samvera.

In order to provide our digital asset repository users a smooth user experience and efficient tools to manage large and complex data sets, we needed to invest heavily on the front end DAM application. For this reason, we decided to adopt Samvera (formerly Hydra), and precisely Sufia, a specific branch of the project geared toward self-deposit.¹⁴ We chose Samvera because of its relative maturity and its thriving community with an impressive workforce that had delivered significant updates to the software at a constant pace. Our reasoning was that it would be more convenient to adopt software that was close enough to our workflows, which we would adapt to better fit our use cases, than to start a bespoke, development-heavy DAMS from scratch.

However, this adaptation task required more effort than expected, in spite of having up to 1.25 FTE allotted for this project alone. Even though progress was made, LAKE went into production and releases kept being cut regularly after that, our issue backlog did not decrease. Some improvements, which we regarded as critical but required deep structural changes to the stack, were too large for us to undertake and needed large-scale coordination with several other members of the community, which as a whole was not able to prioritize them high enough even though they acknowledged their importance.

While the Samvera community was extremely helpful with our issues and genuinely interested in them, our use cases departed too far from the general scope of the project for us to adopt Sufia with only slight modifications. Additionally, when Sufia merged into the Hyrax project¹⁵ and all

¹⁴ Sufia was at the time of our adoption the base for ScholarSphere (<https://scholarsphere.psu.edu/about> Accessed on 15-02-2019), the Pennsylvania State University Library digital repository management system.

¹⁵ Sufia merged into Hyrax (<http://hyr.ax/about/> Accessed on 15-02-2019), a project with a similar scope and sustained by a broader academic community, between 2014 and 2015.

development efforts went into this new branch, we found that migrating to Hyrax would be a very demanding task, given the customization that we had made to the code base.

On the one hand, it was difficult to present our stakeholders with a community-sustained product, which needed significant additional development, rather than developing a product from scratch, which is expected to need some time before it is ready for use.

On the other hand, we adopted Sufia very early in its Fedora 4 implementation. Still partly in our self-sufficient mindset and constrained by deadlines, we developed most of the tools we needed ourselves, while the framework was evolving in a way that could have made some of this custom development unnecessary, if we had been more informed on its direction and presented our use cases to parts of the organization more focused on its development strategy. It is possible that investing more in community road-mapping (which was challenging since our internal road map required a full focus of all management resources) might have avoided us a significant amount of bespoke work.

It is hard to tell whether building a front end system from scratch would have been more or less sustainable, but the difference would have probably not been very large. That is likely where a hack box (ideally later to become a separate community project with a more specific adopter base) would have allowed us the freedom necessary to be more responsive to our stakeholders.

In either case, during this experience, we learned the hard way about some deeper operational and conceptual divergences between libraries and museums. These divergences have a reason to exist and should be accounted for in designing tools and methodologies.

With the adoption of Fedora and Samvera came the acquaintance with a number of aspects that we had initially not accounted for or were not even aware of, such as Linked Data, data preservation, and community-driven development. This was a revealing moment that opened up the spectrum of our focus to topics and whole disciplines that we realized were crucial to the long-term success of such a major project.

The environment we stepped into was largely dominated by academic libraries, with research universities unsurprisingly making up the main driving force behind most of the efforts. This was somewhat intimidating at the beginning, especially with our senior management questioning whether we were equipped to sustain development of a system that other institutions had whole dedicated teams to handle.

As a large museum with a dedication to innovation in technology, and an IT team larger than the average, we felt like AIC was poised to pave the road to best practices for building sustainable core systems in museums. The long-term aspiration for LAKE (which remained unrealized during my tenure at AIC) was to become a community tool that other museums could adopt, and to eventually contribute to.

One of the main lessons learned from interacting with academic libraries was the spirit of sharing information. Open access is something that museums have pursued to some extent very recently; however, today it still remains largely aspirational and/or not effective enough as the technical and conceptual infrastructure is still in heavy development and lacks firm institutional buy-in. Conversely, libraries have a long history of interoperability standards that connect them very efficiently, which led to the recent adoption of standards such as Linked Data by many of them as a natural evolution of technology.

3. Results

As LAKE moved into production, the need for summarizing the experience of its implementation and the strategic plans for its sustenance became a priority. Additionally, broader considerations emerged about the change that LAKE brought about and its untapped potential, especially in regard to publishing information and linking it against information published by other institutions. Such strategies and considerations are laid out in this section.

3.1. Was It Worth It?

The question of whether this risk-taking approach to building a DAMS was worth its yield has been raised at every stage of the LAKE development and especially after its first launch, by colleagues from AIC and peer institutions.

From a numeric standpoint, looking at the period shortly following the live date, the results are not overwhelmingly positive. The budget for such a project was relatively low, roughly corresponding to 1.5 full-time senior engineers' cost for three years (the majority of the resources had been spent migrating and cleaning up the data from legacy systems), and while the resulting system had brought unquestionable improvements to the overall museum workflow, the practical implementation needed much more work before it could become a smooth user experience and reduce operating costs and timing dramatically.

Looking at a projected scenario, however, the advantages of our approach were more apparent. Rather than hiring a company to migrate our data to another system that they owned, and that would leave us no more knowledgeable about the strengths and weaknesses of our model, we did this job ourselves, thereby gaining a full understanding of our data and their potential uses. As described in the black box and hack box section, there are business aspects that an institution does not need to fully own or control, others that it needs to. In-depth insight and ownership of one's data belongs to the latter.

The core philosophy behind LAKE was that building a solid repository system upon transparent and widely adopted standards would serve our institutional interest by creating a firm foundation for our information, which would allow for the creation of useful tools for staff and richer audience engagement tools more efficiently. That would justify our initial costs and AIC users' learning curve.

A transparent and standard-based architecture allowed us to consider every component of the architecture as replaceable, no matter how solid and efficient it was at the moment. Even if we had no intention of moving out of Fedora, the fact that our whole metadata and binary repository could be exported in a machine-readable format made us much more confident about the long-term sustainability of the project. Additionally, most importantly, Fedora has made a point of strength of the fact that it can be replaced, instead of trying to make it harder as "old school" software would do.

3.2. Broader Considerations: The Role of Museums in Supporting Open Knowledge¹⁶

After the LAKE launch, broader and more philosophical considerations emerged along with concrete plans for long-term sustainability of LAKE within AIC; especially about how our approach could be repeated, widely spread, and collaboratively improved among museums.

With the implementation of LAKE we wanted to present a model for other museums to adopt with the goal of making institutions more interoperable in the future. The information held by museums is incredibly interesting and rich with relationships. Furthermore, museums have a very solid tradition of dedication to the visual and narrative aspect of information, and for making complex, profound, sometimes controversial concepts available to very broad audiences. This is a richness that can complement the deeper experience of libraries and archives in knowledge organization.

While the relationships between an institution's archival records, given the right data infrastructure underneath, can build valuable knowledge within a (large) institution's body of information, such knowledge can be greatly augmented by connections with other bodies of

¹⁶ In this paper, the terms "data" (always plural), "information" and "knowledge" are used according to one among various interpretations of the *Data, Information, Knowledge Model* (https://en.wikipedia.org/wiki/DIKW_pyramid#Data,_Information,_Knowledge. Accessed on 03-02-2019). Here, by *data* we intend individual machine-readable statements; by *information*, organized sets of data useful to humans e.g., for drawing conclusions and making decisions; *knowledge* is intended as more complex, non-linear and multi-layered information. It introduces factors of heterogeneity and uncertainty of data sources, a higher degree of abstraction, and ultimately the *Open World Assumption* (https://en.wikipedia.org/wiki/Open-world_assumption. Accessed on 03-02-2019) predicated by Linked Open Data ([1], pp. 83–84).

information maintained by other institutions. The development of a shared communication standard is the first condition to achieve this.

A recurring question which came up, mostly indirectly, during the course of the LAKE implementation, was why our institution would want to invest time and money toward these outward-facing goals, rather than just toward improving its internal production workflows. That seems like a perfectly rational concern from a technical and financial point of view; however, for a so-called memory institution [2,3], the choice of these outward-facing efforts should be judged within the broader picture of the institution's mission and considering the image of the institution's role in the surrounding community that such choice projects.

3.3. *The Neither-Small-Nor-Simple Step from Sharing to Linking*

There are strong motivations for memory institutions to open up their information free of charge to the public [4], including the prospect of tangible gains. A study [5] analyzes established concepts of ownership, control, and revenue generation related to content rights within the Museum sector, as well as the quickly evolving digital communication technology and culture challenging such an establishment. This study concludes that a strategically planned open access policy may lead to

[a] strengthened institutional brand, increased use and dissemination of collections, and increased funding opportunities. ([5], p. 2)

This reasoning is gaining consensus among museums, whose use of image rights for revenue purposes is becoming more and more conflictual with their core educational mission [6] especially in the light of today's technology that enables free circulation of ideas and high-quality contents. Offering high-quality media free of charge and free to use has recently become a sign of progressiveness and inclusiveness. Limitations to this opening are mostly of legal nature, e.g., a legacy of restrictive image rights from artwork owners, and/or technical, such as lack of funds or in-house expertise to digitize and publish large image archives; but in general, museums are increasingly shifting their policies in favor of open access. The frequent presence of opinions on this topic in specialized forums, quite consistently positioned, demonstrates the attention towards it¹⁷.

Such a shift has the positive net effect of more data becoming accessible with a relatively contained effort. However, these data are in many cases only useful within the boundaries defined by the publishing institution's body of information. There is rarely a way to relate to other institutions' information.

This is where museums could and should do more, albeit the step from Open Data to *Linked Open Data* is neither small nor simple, and its perceived advantages may be less appealing.

All along the implementation of LAKE, the Linked Data factor had been regarded by most stakeholders as an incidental feature that came with adopting Fedora rather than a key design strength. At certain times, its advantages had been questioned vis-à-vis the complexity of integrating, storing, and querying this data format. Some of this criticism was understandable, seen from a strictly institutional business-centric perspective, but that need not be the only perspective that a museum's IT should follow.

While the long-term advantages of LOD are proven [7], such advantages are mostly centered around community benefit than institution-specific advantages. On the other hand, the tools to make its adoption simple and within the reach of even large institutions with an established IT workforce, let alone smaller, less funded institutions, are still scarce. Triplestores, the key technology to store and query Linked Data, are still a very narrow niche; this technology is not nearly as stable and efficient as the one of relational databases (e.g., Oracle, MySQL), and their design is more complex than other

¹⁷ Museums and the Web Archive: search for 'open data': <https://www.museweb.net/?s=open+data> Accessed 03-02-2019.

NoSQL stores such as key-value or document stores, which offer a more intuitive and/or familiar data model.

All this means high startup costs for benefits that may not be obvious to the institution as a whole. Since LOD is all about participation (the more, and the more diverse, the merrier), what seems to be the missing factor for LOD's success in the museum world is a "critical mass"¹⁸ that can "bootstrap" a trend of expansion in adoption. This "bootstrap" process should bring down the above mentioned initial costs and complexity, so that institutions with smaller budgets and limited or no IT support can join. At the same time, this critical mass should foster the growth of a culture of participation and collaboration that can become a more compelling mission-related goal, e.g., like open access is becoming today.

Some notable efforts have been, and are being, made to render LOD more understandable and usable by museums, and more broadly, to foster collaborative practices that overcome the initial costs of adopting standards for a long-term gain and to the general public's benefit.

3.4. Collaboration at the Core of Production

The LAKE research and implementation, given its novel character within the museum sector, necessarily projected AIC outwards in search of direction and support. The initial challenges and lack of reference points were counterbalanced by a broadening of our visual field that came with the contact with communities which have collaboration among their core values. Given the amount of effort that these relationships required to be maintained, and the limited resources available to our team, the choice of which relationships to nurture needed to be strategic and built into the project's budget. The following points summarize the key communities dedicated to Linked Open Data or, more broadly, focused on the common development of standards for better interoperability, that we discovered and interacted with along the LAKE development.

3.4.1. Fedora Repository

Fedora has striven since its inception to be a highly flexible and interoperable system. From version 4.0 on, it has adopted the Linked Data Platform (LDP)¹⁹ standard, as well as strict adherence to HTTP and other widely adopted standards. Fedora tries to invent the least new modes of communication in order to focus on being accessible.

In more recent years, an effort has been made separately from software development to define a Fedora API standard²⁰. This standard is built upon LDP, which itself is built upon Linked Data. This allows independent adopters to build their own implementation of Fedora; as long as they follow the defined API standards, all these implementations should behave predictably and consistently at that level. I personally liken this approach to the IIIF specifications described further below, with the exception that Duraspace, the organization supporting Fedora, is committed to supporting one concrete implementation of its specifications as well.

From the Fedora website²¹:

Fedora is a robust, modular, open source repository system for the management and dissemination of digital content. It is especially suited for digital libraries and archives, both for access and preservation [...]

The Fedora project is led by the Fedora Leadership Group and is under the stewardship of the DuraSpace not-for-profit organization providing leadership and innovation for open source technology projects and solutions that focus on durable, persistent access to digital data.

¹⁸ This term was borrowed from <http://americanartcollaborative.org/> (accessed on 29-12-2018).

¹⁹ <https://www.w3.org/TR/ldp-primer/> (accessed on 08-02-2019).

²⁰ <https://fedora.info/2018/11/22/spec/> (accessed on 08-02-2019).

²¹ <https://duraspace.org/fedora/about/> (accessed on 02-15-2019).

In partnership with stakeholder community members DuraSpace has put together global, strategic collaborations to sustain Fedora which is used by more than three hundred institutions. The Fedora project is directly supported with financial and in-kind contributions of development resources through the DuraSpace [values and benefits].

Fedora was at the core of LAKE. It was the project and community that AIC invested most management resources on and that yielded one the most fruitful collaborations. As the only museum in its leadership group, AIC brought some unique use cases and participated in outlining a long-term strategy for the Fedora specs, software, and community.

3.4.2. International Image Interoperability Framework (IIIF)

IIIF is one of the best examples of a technology bringing together Libraries, Archives and Museums. While IIIF does not explicitly mandate the use of RDF or Linked Data, it is conceptually very close to LOD in its end goals, approaches, and audience.

From the IIIF website²²:

Access to image-based resources is fundamental to research, scholarship and the transmission of cultural knowledge. [. . .] Yet much of the Internet's image-based resources are locked up in silos, with access restricted to bespoke, locally built applications.

A growing community of the world's leading research libraries and image repositories have embarked on an effort to collaboratively produce an interoperable technology and community framework for image delivery.

In 2017, about 400 million images have been reported to be available online via IIIF²³; in 2018, a more in-depth census counted over 1 billion IIIF images²⁴. This great success has likely been helped by the relative ease of implementation of IIIF and its straightforward program, which clearly highlights the advantages that it brings, along with the very subject of its focus: images, which every institution today accumulates in great quantity and constitute a common challenge for access and delivery.

IIIF has been implemented in many cultural heritage contexts. Museums have not been as fast adopting it as libraries, but there are some pioneering examples, such as the National Gallery of Art in Washington, DC, and the Yale Center for British Art. A monthly meeting²⁵ gathers museum professionals engaged with IIIF to discuss use cases and showcase new implementations and creative uses of IIIF. Museums have generated among the most creative interpretations of the IIIF protocol and its tools, and this has probably been favored by the richness of museums' visual materials as well as by the importance they give to presentation and end-user experience.

3.4.3. American Art Collaborative

AIC informally joined the American Art Collaborative (AAC) as an "observer" member in 2014. This consortium of memory institutions has the goal of establishing the above-mentioned "critical mass" of cultural heritage institutions and their data to create a cross-institutional information repository. This initiative was very encouraging for AIC which had started moving in the same direction without knowledge of this project. AIC was technically well-poised to collaborate, both for the richness of its collection information, and for the fact that it was already modeled and encoded as Linked Data.

From the AAC website²⁶:

²² <https://iiif.io/about/> (accessed on 29-12-2018).

²³ <https://iiif.io/event/2017/vatican/> (accessed on 29-12-2018).

²⁴ <https://iiif.io/event/2018/edinburgh/> (accessed on 29-12-2018).

²⁵ <https://iiif.io/community/groups/museums/> (accessed on 29-12-2018).

²⁶ <http://americanartcollaborative.org/> (accessed on 29-12-2018).

The American Art Collaborative (AAC) is a consortium of 14 art museums in the United States committed to establishing a critical mass of linked open data (LOD) on the semantic web.

The Collaborative believes that LOD offers rich potential to increase the understanding of art by expanding access to cultural holdings, by deepening research connections for scholars and curators, and by creating public interfaces for students, teachers, and museum visitors. AAC members are committed to learn together about LOD, to identify the best practices for publishing museum data as LOD, and to explore applications that will help scholars, educators, and the public. AAC is committed to sharing best practices, guidelines, and lessons-learned with the broader museum, archives, and library community, building a network of practitioners to contribute quality information about works of art in their collections to the linked open data cloud.

What stands out from looking at the list of museums participating in AAC²⁷, is that not all of them are large, generously-funded institutions. The key of the AAC model is the creation of automation tools to convert the contributing institutions' data sets into RDF, the data format for Linked Data. The conversion itself was also undertaken by collaborators funded by the grant. This way, the only technical burden that participating institutions had to sustain was to select the records that they wanted published.

The goal of AAC is to create the above mentioned "critical mass", i.e., a sufficiently large and articulated network of institutions and their data that can constitute an interesting corpus of knowledge. From this corpus, information coming from different institutions can be discovered, along with relevant cross-institutional relationships. One example would be displaying a Web page about an artwork by author X, owned by Museum A, which also displays links to artwork by the same artist owned by museums B, C, and D. More interesting connections can be built between artworks (e.g., preparatory drawings for a canvas, editions of the same print, etc.), exhibitions, etc. A practical demonstration of this has been showcased for a "browse app"²⁸ which is currently under development.

I have not been able to see AIC join AAC as a full member during my tenure, mostly due to other institutional priorities with more concrete goals taking over. The participation of a large and prestigious institution may draw increased attention to the project and hopefully reach the mentioned critical mass.

3.4.4. Linked Art

Linked Art is a group of cultural heritage professionals dedicated to developing the semantic framework which AAC needs to fulfill its strategic goals. Linked Art is not subordinated to AAC, but the two groups collaborate tightly and share resources.

From the Linked Art website²⁹:

Linked Art is a Community working together to create a shared Model based on Linked Open Data to describe Art. We then implement that model in Software and use it to provide valuable content. It is under active development and we welcome additional partners and collaborators.

One of the core purposes of Linked Art is to create Linked Open Usable Data (LOUD) [8], meaning that it strives to create a model based on the CIDOC CRM³⁰ which is oriented at practical use. The

²⁷ <http://americanartcollaborative.org/about/members-of-the-american-art-collaborative/> (accessed on 29-12-2018).

²⁸ <http://americanartcollaborative.org/tools/browse-app-demo/> (accessed on 29-12-2018).

²⁹ <https://linked.art/> (accessed on 29-12-2018).

³⁰ <http://www.cidoc-crm.org/> (accessed on 29-12-2018).

CIDOC CRM, due to its great complexity and all-encompassing scope, is better suited for direct use by machines than humans, and for harmonizing heterogeneous concepts from diverse sources rather than modeling an institution's cataloging records; Linked Art's goal is to use a limited set of the most practical aspects of the CIDOC CRM to this purpose.

AIC has participated mostly in a receptive mode in the Linked Art discussion; however, the potential for more significant contribution is great, given the richness and diversity of the AIC collections and the complexity of its information management processes.

3.5. A Small Museum's Take on Linked Data

A living proof of the fact that even smaller institutions, if driven by enough will power and long-sighted, but very concrete, goals, can be active players in communities such as Linked Open Data and IIIF, is represented by the Georgia O'Keeffe Museum in Santa Fe, NM (the O'Keeffe)³¹, a relatively small, thematic museum embracing the challenge of Linked Data and other standards through the collaboration with the American Art Collaborative and other communities.

A particular aspect of the O'Keeffe is that its thematic focus contrasts with the geographically dispersed body of work of the artist it is dedicated to, of which the museum owns only a small part. Linked Data is an ideal way to connect the museum with external collections related to Georgia O'Keeffe's artworks, to link items and testimonials of the artist's life with her many artworks dispersed across many locations.

The following is an exchange with Elizabeth Neely, Curator of Digital Experience at the O'Keeffe Museum, who is one of the main drivers of such collaboration.

SC: How did you get involved with the American Art Collaborative?

EN: The O'Keeffe is not part of the AAC, but we felt like we were better equipped as an institution to take on LOD/IIIF by learning from and having contact with the AAC and its members. We hired Design for Context as our development partners partly due to their relationship and experience with the AAC so that we could maximize chances for interoperability and standards adherence. If everyone did LOD differently, we'd all be sunk in a sea of bespoke mapping complexity!

The O'Keeffe hoped to learn and use developed resources from the AAC both through the Design for Context³² relationship and by tracking closely the work of the AAC itself and specifically we attempted to map to Linked Art as a target model on the LOD. And hope to remain an example of a museum that is not part of the AAC, but learns from and is able to expand the network of LOD/IIIF assets.

SC: What value do you see in contributing to a LOD project? Which contents do you think will be most valuable for your institution and for the other members of the community?

EN: The Georgia O'Keeffe Museum is the largest repository of O'Keeffe's artwork, personal effects, and related archives, including important oral histories, correspondence, ephemera, and photographs, as well as a fine art collection of her work, two historic homes, and other accumulated resources that provide rich context for understanding her art, life, and times. The Museum has a diverse set of collections with many potential links, providing insights across the varied holdings and potentially outside, to related linked data resources from other institutions. At a basic level, every data point in some way relates back to the artist. Thus, while the O'Keeffe is a relatively small museum to take on a linked open data project, its discrete and manageable set of information puts it in a unique position to work towards a platform for cross-disciplinary study for other cultural heritage institutions.

Choosing to use a semantic linked data model adhering to museum standards, some well-established and others emerging, made the project in some ways more complex than initially envisioned. As a small/medium museum, the team felt that it was important to show that museums

³¹ <https://www.okeeffemuseum.org/> (accessed on 02-08-2019).

³² <https://www.designforcontext.com/> (accessed on 02-08-2019).

of our scale could participate in establishing appropriate standards and that the benefits would be substantial. By thinking of our project within the field-wide context of collections data access, we realized that we were making it possible to not just intermingle data within this museum - but to prepare data to be shared and accessible as part of a larger network of museum and non-museum data throughout the world.

SC: Which challenges have you encountered or anticipated?

EN: For systems to work together, people need to work together—These efforts affected several departments and a reflection on how we were managing all of our data and data systems. This initiated collaboration far beyond just what this LOD project had as an outcome. The museum has since re-organized to use de-silo and implement team based project planning. I wouldn't call this a bad challenge - but it is an inevitable aspect of a successful project like this: everyone now works differently than they did before (and that's good.)

Also, it was tricky to plan around a community-based project such as the AAC. We had expected certain problems would be solved by the AAC work, but when they weren't (due to priorities, time, etc.), we needed to fold that extra work into our plans which affects time and budget (and can't be avoided.)

Finally, every longer term project has issues with staff turnover - both internally and with our development partners. This again affected time and budget and was a challenge.

SC: How was the idea received by your museum? Which departments were most on board and which ones did you have most questions/challenges from?

EN: Everyone is on board, thought it did make for more substantial changes in work practices in certain departments. It took a while for everyone to see what the potential is behind all of their efforts and practice changes. In fact, I still make little pilot projects (like our data visualization) to help everyone see what they've been part of. That is why we also wanted the beta of collections—it's where the more experiential parts of the team start to get their 'aha moments'. The effects of the above challenges pushed back the parts where the whole team could get their hands on the data in ways that they could understand—it's very important to not push that back too far. I have plans to continue this prototyping and helping staff (and others) understand the benefits of what we're doing.

SC: Which other kinds of collaborative projects is your museum engaging with?

EN: I have lots of plans—not sure where to start! But I do want to think about using our new platforms for digital scholarly publishing that allows for data inquiry and narrative—I'll be looking for people to think about that with us!

4. Conclusions

The LAKE implementation path, as presented in this paper in a narrative form, highlights the evolution of the developing team's mindset triggered by the challenges that emerged along the way and their resolution. The outcome of this implementation was not only a defined product, but also a gain in institutional experience about the subject and an improvement of our ability to sustain the long-term evolution of the product. This experience emphasizes a collaborative approach to problems and attempts to place museums in a specific position within the Digital Humanities discourse, by finding their unique traits, as well as the many points in common with libraries and archives.

Museums are uniquely positioned to contribute to the Digital Humanities community and benefit from it. There are many aspects of the knowledge they hold, their attention for presentation and their skills in data visualization, that reveal a great potential for creating a synergy with other memory institutions. Many technologists are aware of this potential and are making very courageous steps toward that synergy.

What is needed is more buy-in from museums' senior management and a better technological framework to lower the costs of participating in this community. Small, thematic museums are as important as large encyclopedic museums in this landscape, but in reality, almost only the former and only in exceptional cases are able to contribute. It may be mostly the larger, more influential museums

who should initially take charge of paving the way to a more accessible participation and a more inclusive community. This community should create a culture that regards the adoption of standards and cross-institutional information sharing as a mission-related priority.

There are precedents to draw from: Open-access, high-quality images are slowly gaining popularity among museums, and this practice is proving its advantages. In a broader context, OSS is at the core of the business model of many organizations, both non-profit and for-profit.

One of the key factors of these precedents is trust. The inherent transparency of OSS makes it very hard to hide faulty or ill-intentioned design, thereby making the producing company and/or community trustworthy and improvement easier. Public domain contents eliminate the uncertainty and complications of rights management; therefore, one is not worried about getting sued for making the “wrong” use of some content. At the same time, an institution can better limit the proliferation of poor quality reproductions of its collections by offering high quality, authoritative reproductions with no strings attached.

There may be other factors for the success of these models, and the efforts to adopt open standards that encourage resource sharing such as Linked Data, the Fedora API protocol and IIIF may learn from them. Encouraging the display of one’s own collections in a collective portal such as AAC’s or in a student’s weekend experiment, rather than limiting their presentation to institution-controlled websites or mobile apps, could project an institution’s image of openness, inclusiveness, and good citizenship. Such principles, if a culture encouraging their importance exists, could in the long run outweigh the challenges and costs associated with these changes.

Funding: This research received no external funding.

Acknowledgments: The author would like to thank Elizabeth Neely for her prompt responses in the interview on the Georgia O’Keeffe Museum.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Eero Hyvönen. *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*; Morgan & Claypool Publishers: San Rafael, CA, USA, 2012; ISBN 9781608459971.
2. Sainforth, E. From museum to memory institution: The politics of European culture online. *Mus. Soc.* **2017**, *14*, 323–337. Available online: <https://journals.le.ac.uk/ojs1/index.php/mas/article/view/646/604> (accessed on 16 February 2019). [CrossRef]
3. Hjørland, B. Documents, Memory Institutions and Information Science. 2000. Available online: http://www.researchgate.net/profile/Birger_Hjorland/publication/235287518_Documents_memory_institutions_and_information_science/links/00b49539471b136269000000.pdf (accessed on 3 February 2019).
4. Verwayen, E.; Arnoldus, M.; Kaufman, P.B. The Problem of the Yellow Milkmaid. 2011. Available online: <https://pro.europeana.eu/post/the-problem-of-the-yellow-milkmaid> (accessed on 3 February 2019).
5. Kapsalis, E. The Impact of Open Access on Galleries, Libraries, Museums, & Archives. 2016. Available online: <http://s.si.edu/openSI> (accessed on 3 February 2019).
6. Crews, K.D. Museum Policies and Art Images: Conflicting Objectives and Copyright Overreaching. *Intellect. Prop. Media Entertain. Law J.* **2012**, *22*, 795. Available online: <https://ssrn.com/abstract=2120210> (accessed on 3 February 2019).
7. Oomen, J.; Baltussen, L.B. Sharing Cultural Heritage the Linked Open Data Way: Why You Should Sign Up. Available online: https://www.museumsandtheweb.com/mw2012/papers/sharing_cultural_heritage_the_linked_open_data (accessed on 29 December 2018).
8. Sanderson, R. Publishing Linked Open Data: We Did It. Now You Can! In Proceedings of the MCN Conference, Denver CO, USA, 13–16 November 2018.

