

Greater than the sum of its parts: A model for a national collections database

Matt Connors, Macquarie University, Sydney

Abstract

This paper, based on a presentation given at the 2005 ICOM UMAC conference, suggests a way that the ever-increasing electronic catalogues of individual museums can be discovered by remote researchers and delivered in a way that is meaningful to the researching and ongoing safeguarding of the physical objects. I examine the motivation to create a National Collections Database, and a model for bringing scholarly material out of museum collections, and communicating it with researchers. The examination includes a brief review of Australian efforts so far, and the metadata schema that have been promoted both locally and internationally (such as the Distributed National Collection and the Open Archives Initiative Protocol for Metadata Harvesting). The model discusses ways in which authorised researchers can query participating museums about its holdings, and for research-quality data to be returned, and illustrates participation methods suitable for small and large museums. As the model's greatest potential is the capacity to minimise the gap between museums and that traditionally indispensable research tool: the library, I examine systems that are almost taken for granted in the library environment, systems that could allow museums to be viewed similarly as an essential service in higher education.

My presentation describes a model of a National Collections Database to connect museum users to bring museum object records, and do so at a level of detail that is valuable for research and academic purposes. Most crucially, the model also supplying researchers, primarily familiar with library systems, who may not have discovered the content of research-focused museums. It is about moving beyond the current web presence of museums, by allowing them to offer more than stand-alone web sites and online catalogue samples.

Australian museums are being encouraged to bring their collections online. Note projects offered jointly to the Australian museum community, such as *Capture your Collections*,¹ a publication produced in collaboration between Collections Australia Network (formally AMOL) and the Canadian Heritage

¹ *Capture your Collections* is available to museum professionals from the AMOL website. It describes procedures for digitising, mainly images of small collections. <http://amol.org.au/capture/>, May 2002.

Information Network. This resource offers a start-to-finish solution for digitising images and adding them to your collection.

A primary resource for the museum community, CAN has made a significant start on the concept of a national approach to databasing. CAN's initiation of discussion about museum information reflect this semi-centralised approach to information management, such as the Open Archives Initiative Protocol. Aimed at practitioners of IM in museums and galleries, OAI-PMH is a very simple communication system when compared competing standards. OAI-PMH is all about allowing repositories to communicate with one another. As such, it has little focus on the museum user, as it is about machine-to-machine communication.

There are important steps yet to be explored in OAI-PMH fulfilling the role of bringing national collections together. There is nothing in OAI-PMH, either a manual or automatic system, which allows for resource discovery.

As there is no hierarchy given to the structure of the repositories, each is permitted to harvest from one another. For example, the Australian War Memorial (AWM) may harvest from the National Museum of Australia (NMA), and *vice versa*. But supposing 50 new records become available at the AWM, the NMA harvests their metadata and gives the records 50 new NMA record identifiers. The AWM notices that 50 new NMA records have appeared, and harvests them back, and so on.

Another difficulty with repository discovery arises with record 'identifiers', which are a unique part of an object record for use within the holding institution. Take our 50 example objects: these are given an identifier when accessioned into the AWM. They are given another by the NMA, and another by any organisation that harvests from the NMA. Libraries avoid this problem by identifying the object according to other objects of the same type, rather than where it came from, such as the Dewey Decimal or Library of Congress classification systems. Whilst museums cannot rely upon a central system to

identify unique objects, they can take into account all the data elements applied to the object by another institution.

These issues are not ones that become apparent after a certain scale is reached, but are encountered from the very first two museums joined in this way. This issue will need to be addressed more thoroughly as the development of the OAI-PMH protocol moves away from discussion, and further into implementation.

The OAI-PMH documentation, like the protocol itself, is kept as simple as possible. A benefit of this approach is that it captures many institutions interested in pursuing interoperability. From a user point of view, there are difficulties. A user will not know where to go for the latest information, that is: the most recently and comprehensively harvested metadata. Do they go to the NMA, or the AWM? Who has the most recent metadata? Which museum has the most comprehensive, rigorous and active policy of harvesting metadata as widely as possible? What standards are being applied? Most users simply have no way of knowing this information.

“Museum collection databases often do not provide profound intellectual information for the users. “Contextless digital bits” travel all over the world changing the way the original artwork is interpreted”.²

Ideally, an information management system should include input from information architects, and take advantage of resources currently available only to large museums, but always maintain a clear focus on the benefits to the end user.

The architecture of a 'fully-integrated' IM system could be viewed as the best existing system in several ways. Figure 1 is a schematic model of the MCM Database at Macquarie University. This was introduced to make a record of

² V. Kravchyna and S.K. Hastings, 2001, *Informational Value of Museum Websites* http://www.firstmonday.org/issues/issue7_2/kravchyna/, Dec 2001

over 200 000 moveable heritage objects at the University.³ It is a fully-integrated system, which means it combines the various servers it needs into the management application itself. The web server is integrated, so is a Z39.50 server for feeding data to library management systems using MARC metadata schema.⁴ Z39.50 is a protocol designed to transmit the richness of a complete data set to another system in a way capable of understanding it (such as a MARC Library system). These are still machine-to-machine communication protocols.

This system, however, is difficult to scale. It relies on a LAN connection for content management, meaning web users are limited to data-out only, at least until an open-source solution presents itself.

Figure 2 is an architectural schematic of the NCD model. To remain technology-neutral as far as possible, I have not explored in detail connection types that would be appropriate, except to say that, because the user interface is delivered over the Internet, HTML, SQML and XML (types of communication protocols commonly used on the Internet) would be the most accessible connection modes.

It is the role in the model of museum IT professionals to establish communication protocols that satisfy their own data security needs. Whether this involves firewalls or similar, is not part of the model architecture.⁵ The exception to this is an authentication server, because it changes the model architecture.

Metadata is drawn from existing museum IM systems and housed centrally. This metadata is used to respond to queries from different user groups, such as public, research and professional users. Their requests return information

³ By way of comparison, the National Trust holds a record of non-moveable cultural heritage objects. <http://www.nationaltrust.org.au/property.htm>, November 2001

⁴ Kody Janney and Jane Sledge, 1995, *A User Model for CIMI Z39.50 Application Profile* http://www.cimi.org/public_docs/Z3950_app_profile_0995.html, May 2001

⁵ There are more sophisticated ways to manage rights than using firewalls, which assume everyone inside a firewall is trusted, and everyone outside not trusted. One such system is being developed by IPR Systems, Australia. <http://www.iprsystems.com>

from the data repository of the museum supplying the metadata, the detail of which depends upon user rights. This is a user-focussed model relying on components from proprietary IM systems, and components from emerging standards and protocols.

Users

Data creators, such as curators and museum accessioning staff have primary access to edit all aspects of their data repository, researchers with less access (but to all collections), and general public, or a user not logged-in, would see information on par with what may be available through the relevant museum's website.

Although not based on log-ins or permissions, this concept is already in operation at sites such as the Movie Image Archive Collection. This site allows the user to view fully digitised objects in the collection, such as archive footage. With many objects in the collection over six gigabytes, the database is not intended for the casual user.⁶

Various institutions could adopt the role of administering the user groups that fall outside the staff of a museum, as required. For example, a university may negotiate researcher privileges with the administrator.

Metadata Repository

Metadata is the key to the NCD model, illustrated in Figure 3. Even in a moderately repository, many records need to be described with a number of roles. For example, 'Frank Lloyd Wright' could appear at the subject of a written article, the maker of an object, or could be the title of a written work. Even a keyword search that uses Boolean operands to narrow search parameters is going to miss out on providing a search result that is possible with metadata, such as an artist's rendition of a Frank Lloyd Wright-designed house that refers only to the house's name.

⁶ The Internet Archive site contains many films of several gigabytes.
<http://www.archive.org/details/movies>, Sep 2005

Data Repositories

The various data repositories may be constructed out of proprietary IM systems already in place in museums. Although the model's metadata repository and 'Data Repository 1' are illustrated separately, there is no reason that these two elements could not be housed on the same hardware in the manner of an integrated system. Data Repository 1 is offered as a service to museums with no IM infrastructure of their own.

Access

Access in the NCD model is predominantly web-based. This is because the majority of data retrieval is performed by the user, who will want convenient access, as well as access to other gateways that are already web based, such as the Australian Libraries Gateway.

Data creation is sometimes difficult *via* the web, due to the speed and security of the connection. For this reason an information manager could find it more convenient to access their museum's database through a LAN. Thus 'Museum 1' is defined inside a LAN.

The 'Small Museum' by contrast shows an information manager operating outside a LAN using an online object management system. In this case, the slower response times are traded off against the 'Small Museum' being freed from a need to maintain an information management infrastructure at all. This means that any museum that can obtain Internet access can be joined to the NCD.

Web-based accessioning would also allow very small museums, ones that may lack any form of an electronic record of their objects, to have access to a nationally-sanctioned information management system. CAN have recently announced an important step down this road, by offering a capability in this area.

Discovering Gateways

There is no easy way to discover the existence of portals or gateways. Supposing another country does have or brings online a national gateway, how would the NCD model become aware of it? A group of gateways at a national level would be a list small enough to be monitored and maintained by human administrators. For databases that require frequent checking and/or repackaging of their metadata stream to be successfully searched, one would need to weigh their usefulness to the Australian user base against the maintenance overhead.

The gateway discovery that libraries use is not a suitable model for museums for two reasons. Firstly, museums are not as organized as libraries, and do not have a well-established system of communication standards in place upon which to grow. Secondly, libraries predominantly deal with other libraries that hold collections in the same written languages. Museum objects do not rely solely upon language for their interpretation and usefulness. For example, a museum in Sydney specializing in colonial architecture could find objects and research of significant relevance in a museum in Chile, whereas two libraries in these cities are less likely to draw the same benefit from a similar interoperability.

The key benefit to remember with centrally stored metadata is that searching for information resources is possible even when access to the resource is temporarily severed. In a sufficiently advanced IT system, the delivery of the resource could even be cached to a time when the connection can be restored.⁷

There are a variety of quality standards and information management policies that can be applied to the NCD model. The ICOM CIDOC *Proposed International Information Standards for Museum Automation* is the best one I have examined.

⁷ This system is employed in at least one peer-to-peer file transfer system: Kazaa Media desktop, produced by Sharman Networks.

Meeting community needs

Some Indigenous communities in Australia have prohibitions on viewing images of the deceased, and some migrant communities have prohibitions on viewing images of particular familial relationships. Metadata can be used to filter museum content to make online museums suitable for use in these particular communities.

Data Security and Reliability

Much of the security of this model relies on the distributed nature of IM. Centralised only where needed, the metadata repository is backed up using existing IM resources of an appropriate body, such as a major museum. This means that if a museum is offline for any reason, the metadata associated with that collection could still be served to a user.

This also means that the responsibility for data reliability remains with those most dependent on it and most able to manage it – the curators and accessioning teams of the museum that has care of the objects. The difficulty of managing the metadata, which is effectively distributed content, can be minimised by mandating certain minimum standards that the metadata must adhere to. If these standards are not met, the museum's objects can be restricted from being searchable from the public user interface, until the standard of the data has been rectified.

Rights Management

Rights management is the process through which users are given or denied access to information in the database. An authentication server, illustrated in Figure 4, is one solution that could be applied to the NCD model. An authentication server can handle requests for authorisation and access. Note also that the server can only authenticate users who have been given permissions to use repositories by the repository holders.

Images

Image capture is a disturbing process to a museum object, and often one that an object has to endure repeatedly as new images are often gathered for

different roles of the image, or indeed the changing use of the object, such as research, marketing, insurance and promotion. With a set of guidelines in place, and a reliable location in which to house them (such as the NCD), an object can be disturbed less frequently by repurposing a base set of images that relate *via* metadata to the role of the user.

Making meaning

What the database will not do is take the place of the online exhibition. The NCD model is neutral to the object's context, and applies no interpretation on the objects it describes. Museum professionals in the areas of education and online exhibition design would need to apply their skills to present the material effectively to a 'general visitor' audience. Consider by way of parallel that visitor services / exhibition design and cataloguing / information management are typically treated as distinct categories in the physical museum space.

Object security

Possibly more so than other countries, Australia needs to examine whether digitisation of museum object data is a way to allow the return of the objects to the cultures that created them. One of the arguments against returning Indigenous cultural objects held in museums is that they are lost to the wider community. If the object was digitized before its return in a way that met a national standard, then some levels of research could be performed with the object record alone, such as typology, as well as the record being available to a national audience for the purpose of education. This can also lower the chances of a returned object becoming unlocatable in the future. For example, Object ID is an international standard for describing cultural objects. It has been developed through the collaboration of the museum community, police and customs agencies, the art trade, insurance industry, and valuers of art and antiques.⁸ Such a standard, especially with the credibility lent it by being initiated by the J Paul Getty Trust, could be easily applied to the NCD, making a control point for all objects moving through museums in Australia.

⁸ <http://www.object-id.com/index.html> , July 2002

I appreciate that this article has covered a wide range of concerns in varying level of detail, but I believe that a comprehensive, user-focussed approach is lacking from much of the research in this area of museum practice. To come back to my library comparison, a typical research library user does not want only the contents of the building they are in. This model of museum data usage makes the same assumptions about the users we are used to receiving in our museums. Centralised only where needed, and distributed where possible, this model's greatest potential is the capacity to minimise the gap between museums and traditional research tools, such as libraries.

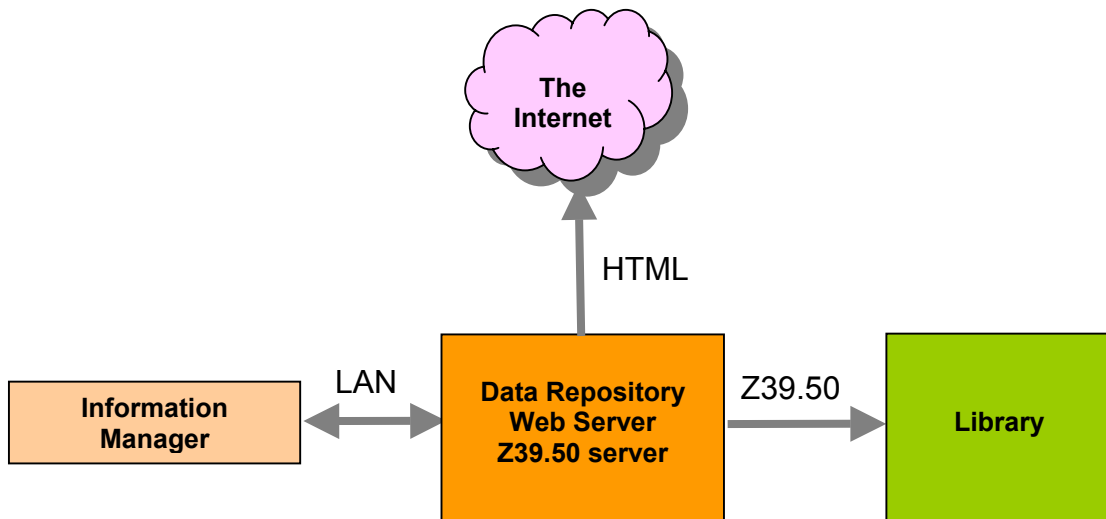


Figure 1: A fully-integrated museum IM system

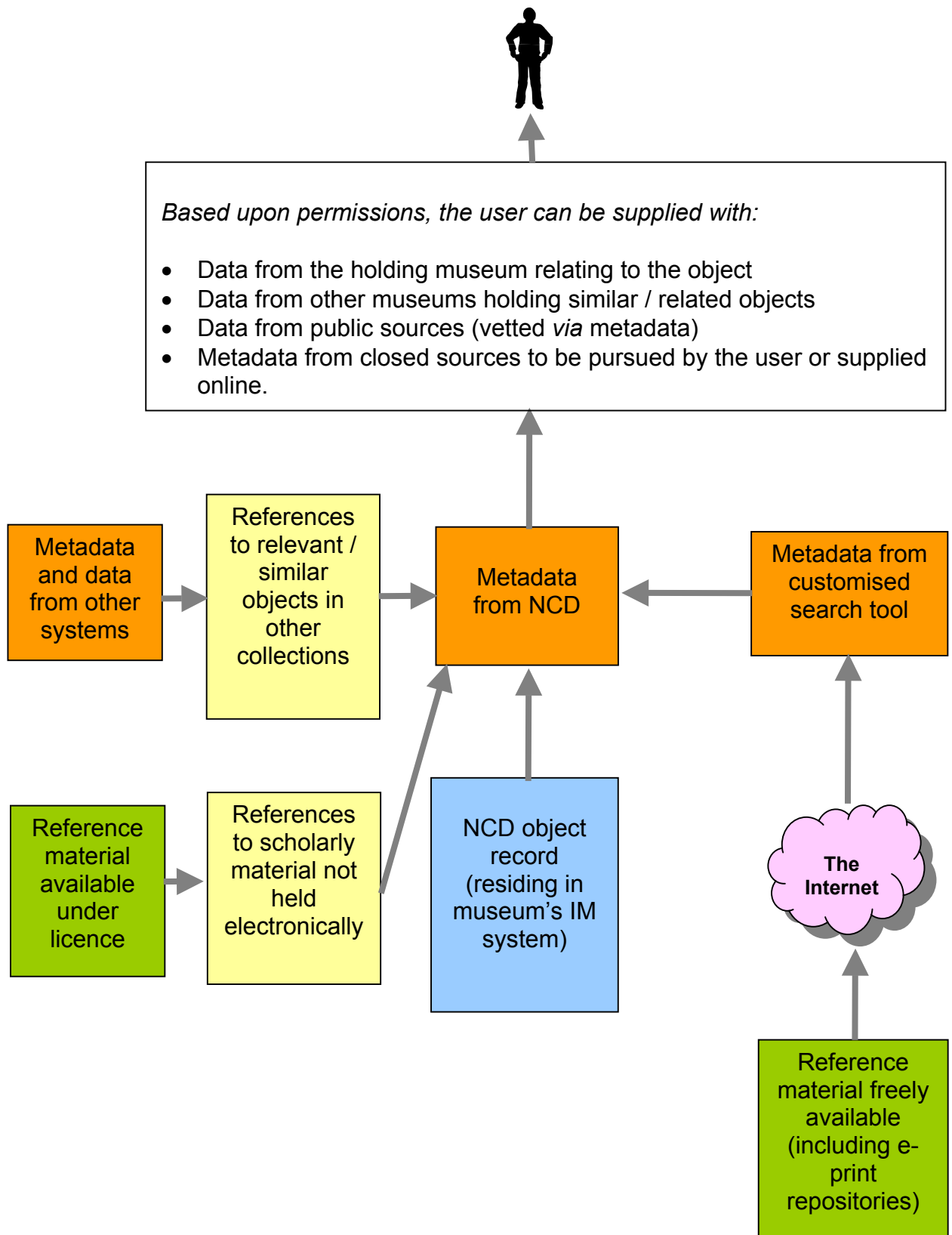


Figure 2: Data sources

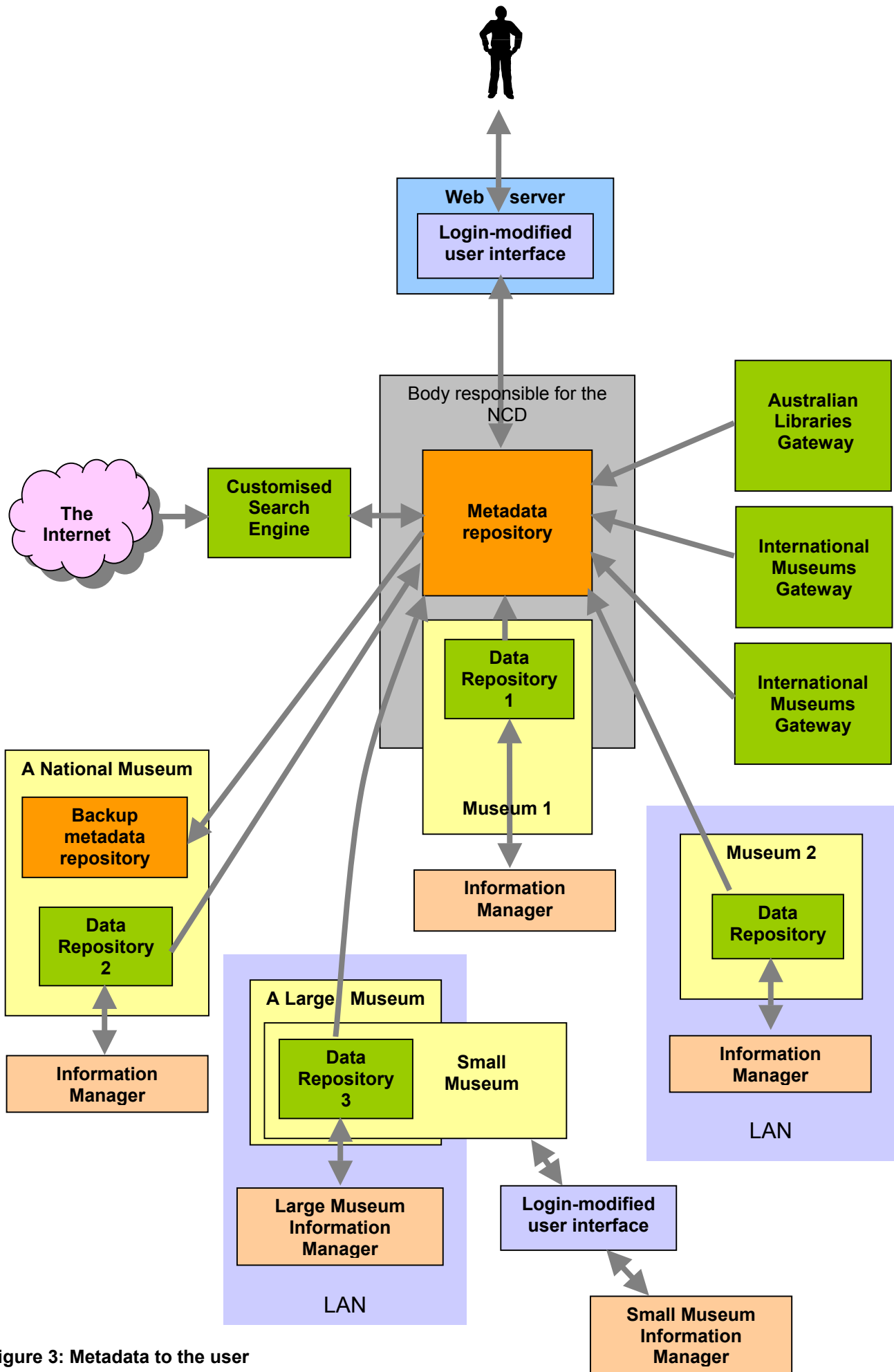


Figure 3: Metadata to the user

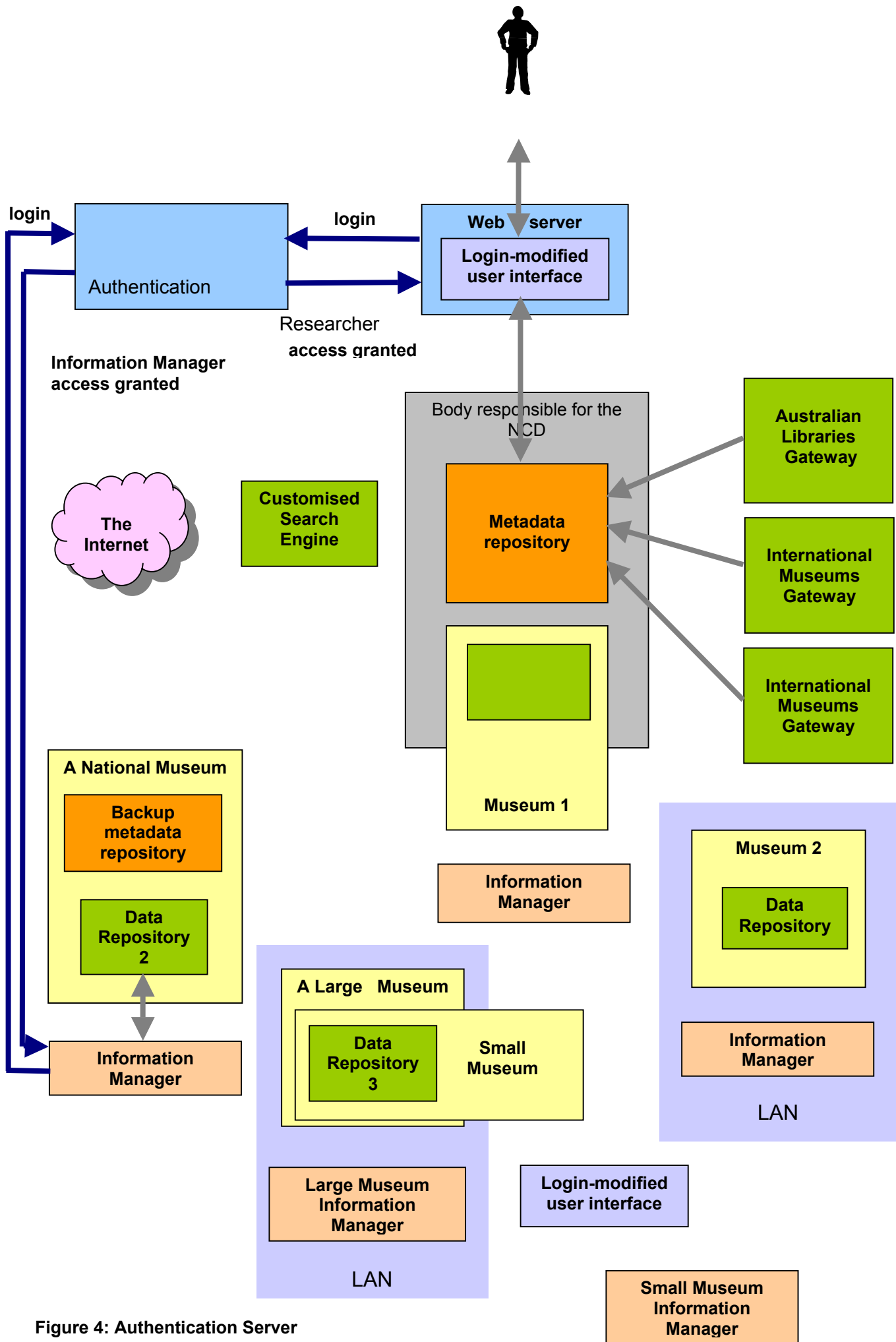


Figure 4: Authentication Server

