# MetRe: Supporting the Metadata Revision Process

Introduction

MetRe is a prototype interface and service designed to support the metadata revision process. Improving the consistency of metadata records within an environment is a common repository management task, not only because of the potential for user error when submitting, but also because of other sources of error, such as systematic error resulting from the limitations of the deposit process chosen. Evidence to support the metadata correction process may be provided by automated metadata extraction tools, evidence from within the repository, or by comparison with best practice across the repository landscape.

Functionality

This prototype is able to identify several characteristic classes of error, twinned with an interface able to highlight several types of individual and systematic error, including a notion of local and general best practice. These may be used in a variety of ways – for ranking in priority order, for browsing or correcting systematic error according to symptom, or directly editing a single record via an AJAX-based customised form. The approach makes use of statistical information gleaned from local (intra-repository) and global (inter-repository) harvested metadata to identify patterns, rank occurrences and co-occurrences.
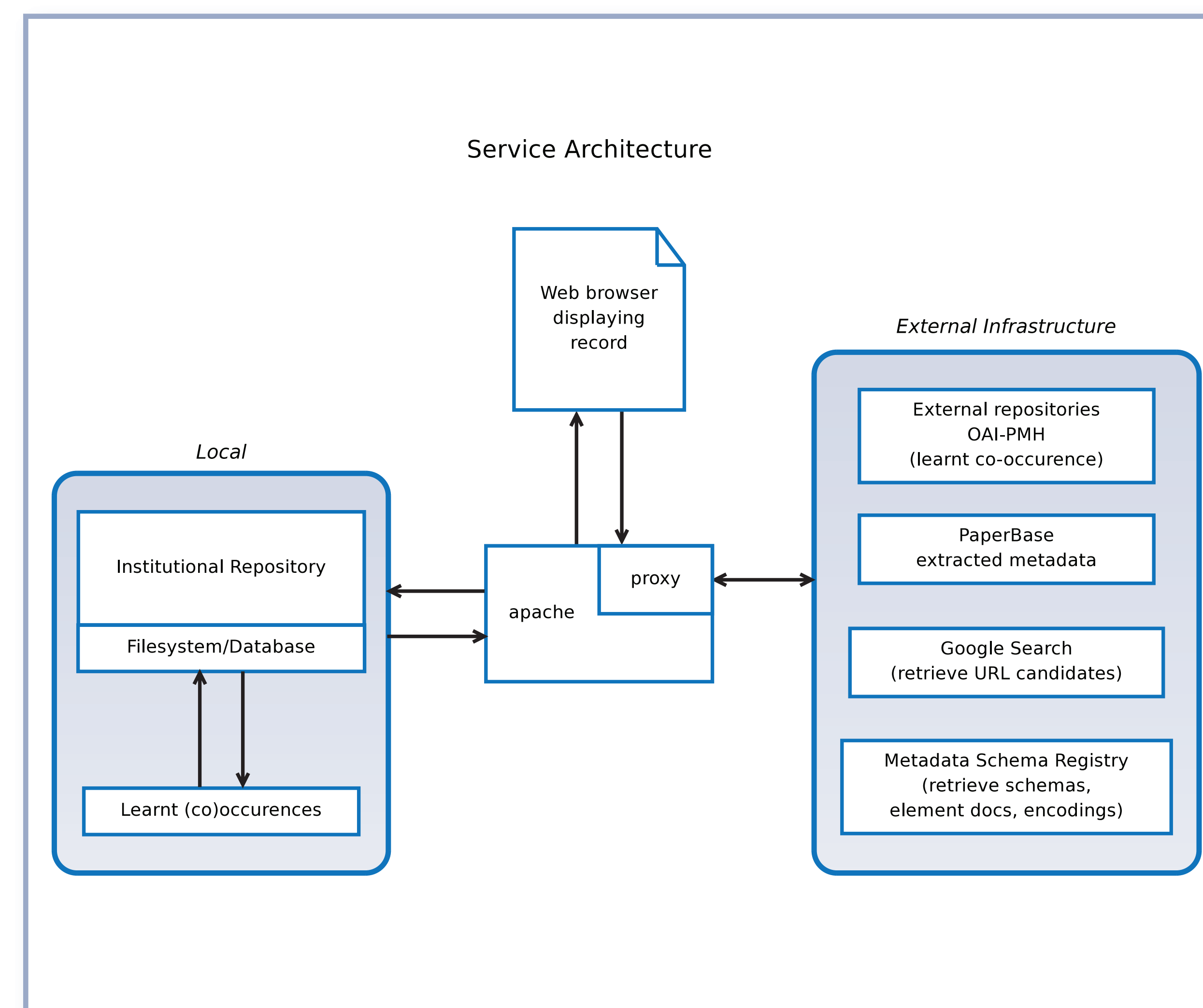
Service architecture for the MetRe prototype

DC Field	Value
dc.contributor.author	William Pearson
dc.contributor.author	Mike Fraser
dc.date.accessed	2006-04-04T16:17:37Z
dc.date.available	2006-06-02T16:17:37Z
dc.date.issued	2006
dc.description.abstract	
dc.identifier.citation	Pearson, W., Fraser, M.(2006).Exploring the surface: discovering differences in Vision-Neutral H
dc.identifier.uri	http://www.cs.bris.ac.uk/Publications/pub_master.jsp?id=2000584
dc.language.iso	en
dc.relation.isreferencedby	
dc.title	Exploring the surface: finding differences in Vision-Neutral Haptic Search.
dc.type	Text
dc.type.publication	refereed

> Title extracted from paper differs
> Extracted title: *Exploring the surface:* **Discovering**
> *differences in Vision-Neutral Haptic Search*
> Recommendation: replace title
> Click on the text to make the recommended change

dc.description.abstract	learning to bootstrap from standard sequence similarity search methods.First a standar high similarity to the target(assumed homologues) and not true for general sequences, zone.To learn the rules HI describes the sequences in a novel way based on a bioinforn logic programming.To evaluate HI we used the PDB40D benchmark which lists seque HI methodly with PSI-BLAST alone and found HI performed significantly better.In a showed that these improvements were robust for all reasonable error costs.The predict provide insight into conserved features of homologous protein families.Conclusions: H central bioinformatic problem.HI with PSI-BLAST is shown to outperform PSI-BLAS obtained using HI with any sequence similarity method.
dc.description.sponsorship	Andreas Karwath and Ross D. King were supported by the EPSRC grant GR/L62849.
dc.identifier.citation	Karwath, A. King, R. Homology induction: the use of machine learning to improve se
dc.identifier.uri	http://hdl.handle.net/2160/258
dc.language.iso	en
dc.publisher	BioMed Central
dc.relation.isreferencedby	http://www.biomed.central.com/1471-2105/3/11
dc.title	prove sequence similarity sear
dc.type	
dc.type.publica	

> Recommendation: On your server, this publisher (BioMed) is usually accompanied with the following rights statement: **The full text is made available here under the terms of the BioMed Central Open Access license.**
> Click on the text to make the recommended change

Exploring metadata consistency via the MetRe prototype

The role of the metadata registry

Pragmatically, the development of 'spelling, grammar and consistency' checkers for metadata is rendered simpler by the existence of machine-readable entity, metadata element and structural definitions, as well as general-purpose service APIs for 'sanity-checking' against other sources. These may include anything from information available on the Web to formal metadata to be extracted from the data object itself, in the event that this is practically achievable. It is therefore in part the increasing availability of such resources as easy-to-access Web services that makes services employing this information relatively practical to develop and to maintain. In the case of MetRe, the IEMSR metadata schema registry is a core component.

Evaluation.

We chose to evaluate the prototype on an intra-repository task, primarily because we felt that this would require less knowledge of general metadata practice, and would allow us to work within a familiar interface and environment as well as a familiar application profile. We built the prototype interface by customising a DSpace page via JavaScript, which interfaces to the MetRe service via a local proxy. Initial user testing of the prototype shows that it is quick and simple to use by comparison to manual editing without assistance, which requires a good knowledge of Dublin Core metadata and reveals itself to be a frustrating task.

Future work.

In future, we plan to:

- develop the system further,
- complete more user testing in a production environment,
- increase the number of types of error or warning that are recognised,
- to make future tools available as a publicly accessible service for general use.

This work will be continued within the JISC-funded FixRep project, which will explore integration of a variety of services and methods into a basic framework similar to that described here.

Emma Tonkin
UKOLN, University of Bath, UK
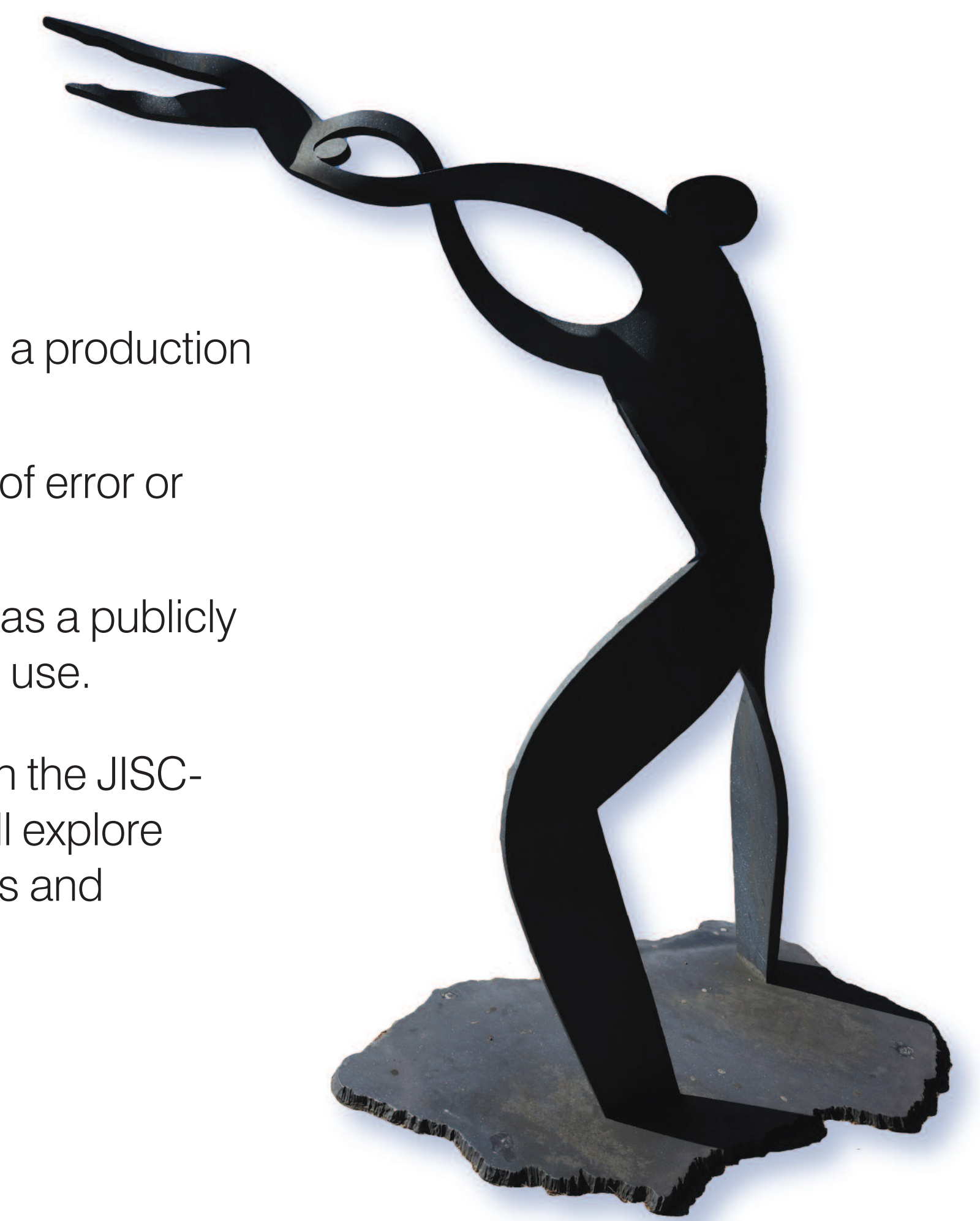
Consistency

"Use of recommended or mandated structures, schemes or policies which might enable software designers and service creators to predict with some certainty the type of content to expect."
– Definition (Charlesworth et al, 2008)

"Inconsistency is a fact of life, and any repository instance or system that wants to avoid bottlenecks is going to have to accept items that have inconsistent metadata [...] That doesn't mean you have to settle for it, though. It's possible to take a progressive approach, where messy metadata comes in, and is then brought into consistency with particular standards."
(Charlesworth et al, 2008)

References

T. R. Bruce and D. Hillmann. The continuum of metadata quality: Defining, expressing, exploiting. In Metadata in Practice, pages 248–249. D.H.a.E. Westbrooks, Editors, 2004.

A. Charlesworth, N. Ferguson, E. L. Morgan, S. Schmoller, N. Smith, and D. Zeitlyn. Feasibility study into approaches to improve the consistency with which repositories share material. Final report for JISC, November 2008

L. Figueredo and C. K. Varnhagen. Spelling and grammar checkers: are they intrusive? British Journal of Educational Technology, 37(5):721–732, 2006. J. Greenberg, K. Spurgin, and A. Crystal.

Functionalities for automatic metadata generation applications: a survey of metadata experts' opinions. Int J. Metadata, Semantics and Ontologies 2006, 1(1), 2006.

L. Swartz. Why People Hate the Paperclip: Labels, Appearance, Behavior, and Social Responses to User Interface Agents. Master's thesis, Stanford University, California, 2003.

E. Tonkin and H. Muller. Semi automated metadata extraction for preprints archives. In Proceedings of the 8th ACM/IEEE-CS joint conference on digital libraries, 2008.