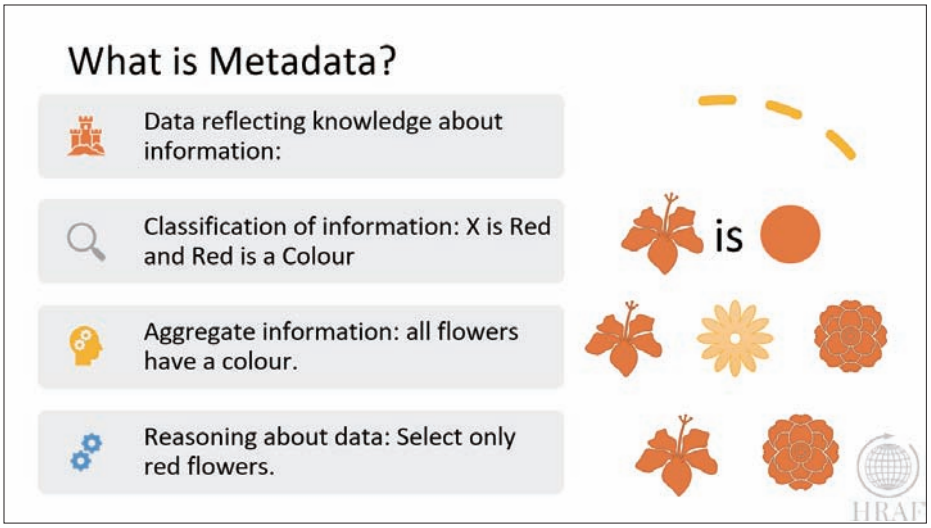
Special Lecture

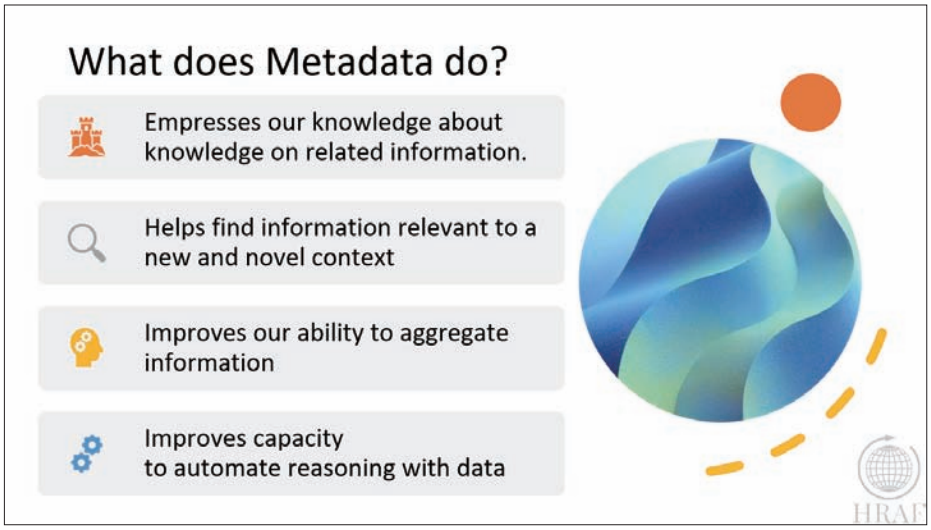
Adapting to the New Era of Intangible Cultural Heritage with Metadata

Michael Fischer
Vice-President of Human Relations Area Files

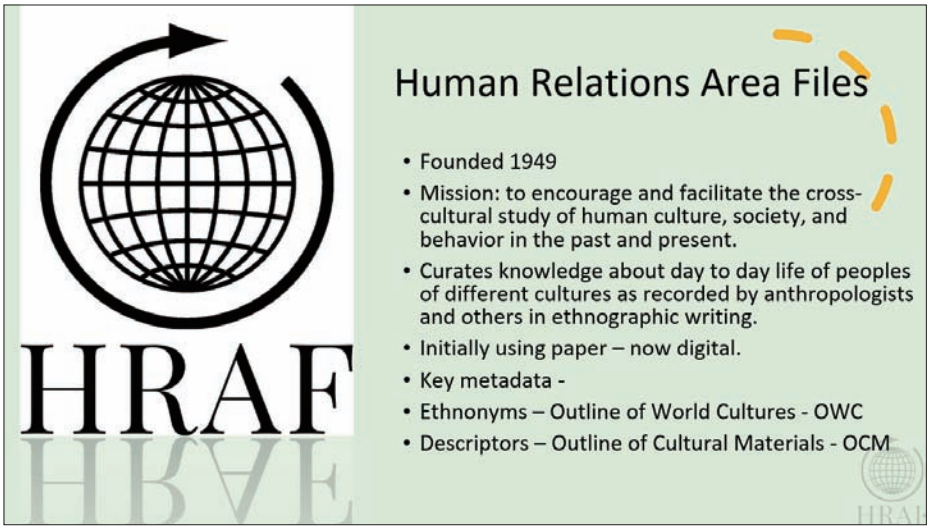
We have been in the midst of a new era of intangible cultural heritage for quite some time. Rapidly vanishing ways of life, globalisation, climate change, political change, new communications technologies, and now a global pandemic, are destroying, modifying and creating cultural patterns, activities and knowledge continuously. Although these processes have been ongoing since the first cultural expressions emerged early in the ancestry of humans, these have accelerated over the past two centuries from gradual generational change to increasingly rapid change over increasingly small sub-generational cohorts. The impacts of these processes have expanded due to major changes in communication, which can spawn and nurture new forms of sociality and shared experience without spatial contiguity. So it has become increasingly imperative to not just collect descriptive and narrative accounts of ICH, but to further document ICH within its context of use and experience.



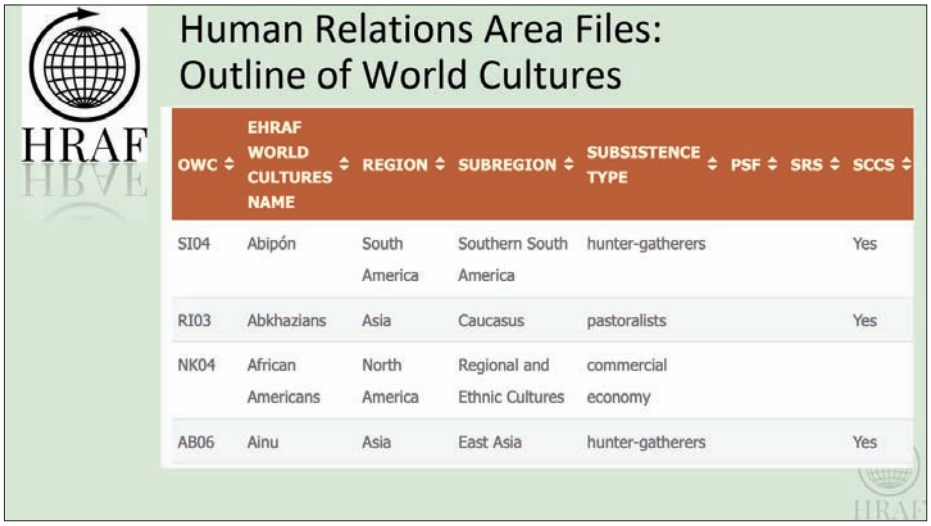
Metadata is a relatively new term (1960s) for a very old human creation. The textbook definition is something like "data about other data" … metadata is documentation that describes data (Research Data Management Service Group, no date). Greenberg's (2003) functional definition of metadata is structured data that supports functions associated with a designated object. Without metadata, a record of behaviour, and even firsthand experience, is not sufficient to capture cultural practices, knowledge and skills. Metadata is often associated with libraries, museums and customer databases. But it is a far more pervasive concept: abstract classification based on perceived qualities of experience with objects and processes. Human language is the main means by which people describe and animate complex and detailed metadata about their lives and the world around them. Language reflects the juncture of bottom up generation of behaviour and presence (phenomena), and top down conceptual categories and relationships which are imposed on phenomena, which often invoke responses that modify the phenomena. The outcome is often an instance of ICH. Without explicit classification learning is difficult, and for many human domains, impossible.



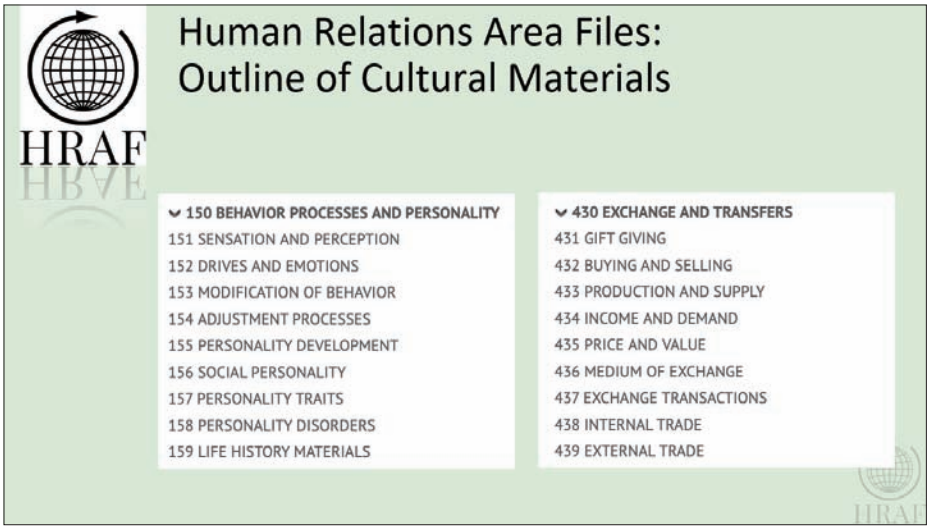
Centres of Tangible Cultural Heritage, such as museums or libraries, have metadata that is often oriented to describing and classifying individual objects - the abstract archetypical classification is often more or less metadata, a connecting thread for a set of objects. In documenting ICH the archetype is the focal unit, and specific instances are given as examples that document the archetype, rather than the focus of attention. Specific, instantiated, examples are needed, but these often serve more as a kind of metadata relating to the generic ICH these are an instance of. Where Tangible Cultural Heritage is concerned with things, and the processes of their making is secondary, ICH focuses on DOING things, ethnological processes, and specific instances are examples of ethnological archetypes. Thus metadata in ICH is largely oriented to documenting processes for making things, whether tangible, such as a basket, or intangible such as a marriage. This is more so in the digital age we now find ourselves inhabiting. Understanding tangible artefacts such as a collection of emails rely on understanding the cultural processes these are a product of.



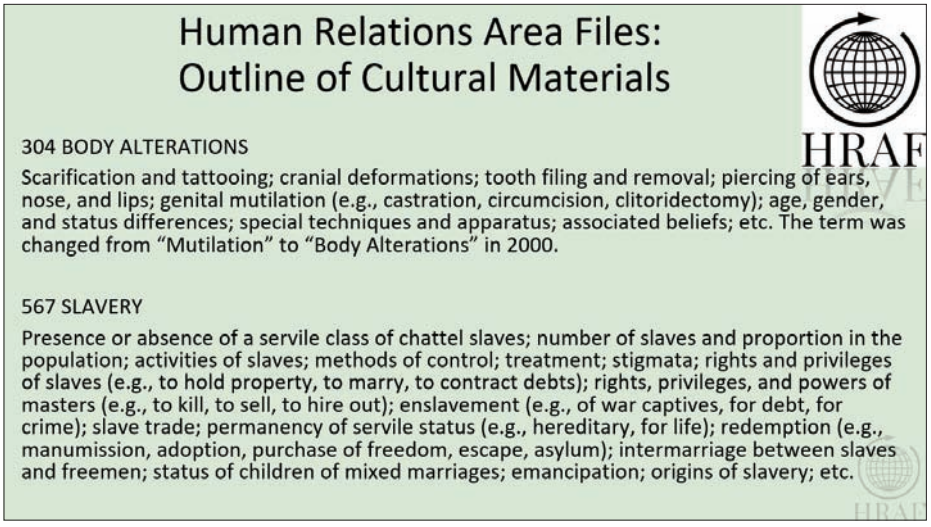
The Human Relations Area Files, and its parent, Yale's Institute of Human Relations (IHR), was one of the first attempts to curate ICH on a large scale. The IHR was established in 1929, and created the Cross-Cultural Survey. In 1949 HRAF was founded to promote the Cross-Cultural Survey on a broader scale. Currently the digital version of HRAF curates over 7000 ethnographic documents covering 360 societies, with a total of around a million pages. Since its inception as the Cross-Cultural Survey in 1935, metadata has been applied to each page and paragraph by professional HRAF analysts. HRAF, like so many organisations sometimes referenced as 'memory institutions', is focused on reuse of information, not just its preservation.



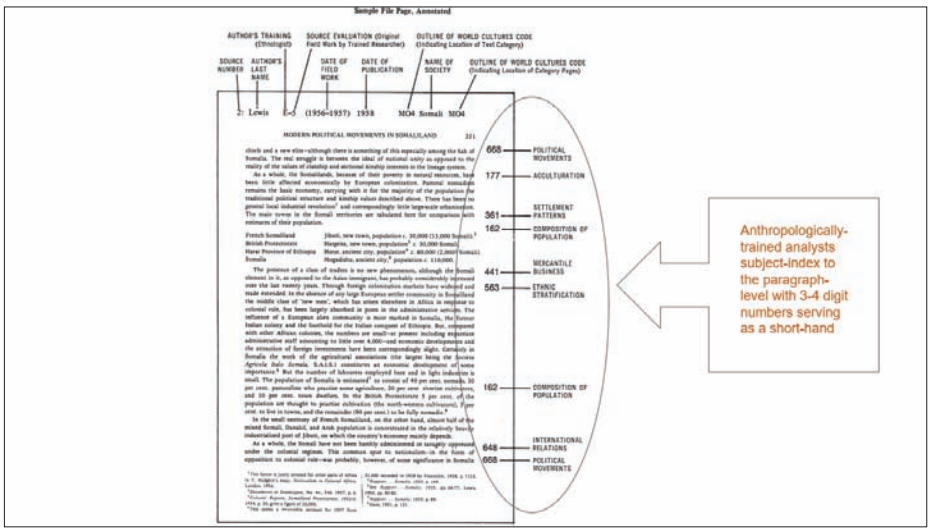
The Outline of World Cultures provides critical metadata for identifying the group associated with the ethnography. Many groups have many different names, and the OWC normalises these to a standard 'preferred' name. These have changed over the years, often in response to indigenous groups requests. But all the ethnonyms are retained as these are used in various documents HRAF curates. The OWC covers many more groups than the 360 societies HRAF curates, so may be of value to other providers. We will be releasing a public OWC service in the near future.



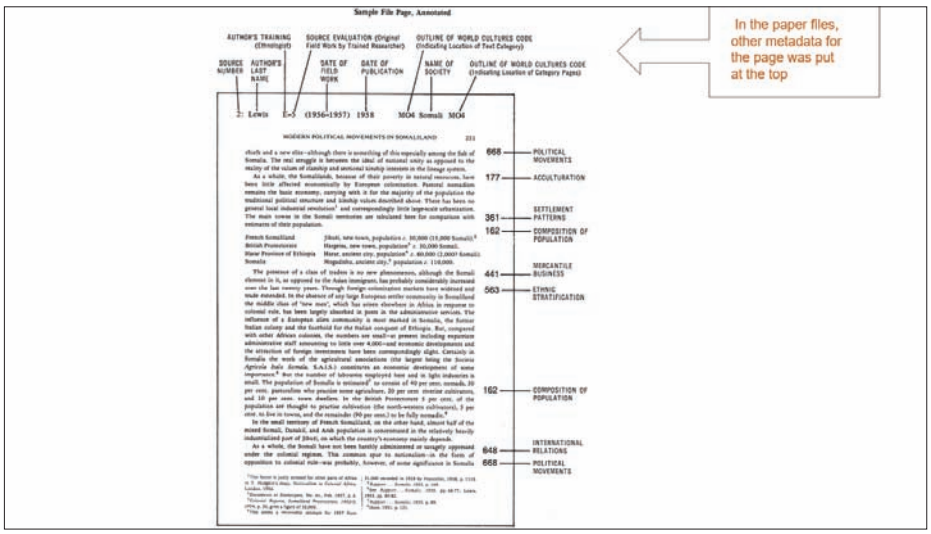
The other critical metadata for HRAF collection is the Outline of Cultural Materials (OCM), originally a thesaurus developed for the the Cross-Cultural Survey with 755 terms organised into groups. Each is assigned a numeric label, and it is this which is applied by HRAF analysts to the text on a paragraph by paragraph basis. As an attribute of the text, the OCM indicates what kinds of ICH processes a given text refers to.



Each of the terms in the OCM is defined. These definitions have changed occasionally to reflect a more detailed understanding that has arisen over the past 85 years since the inception of the OCM.

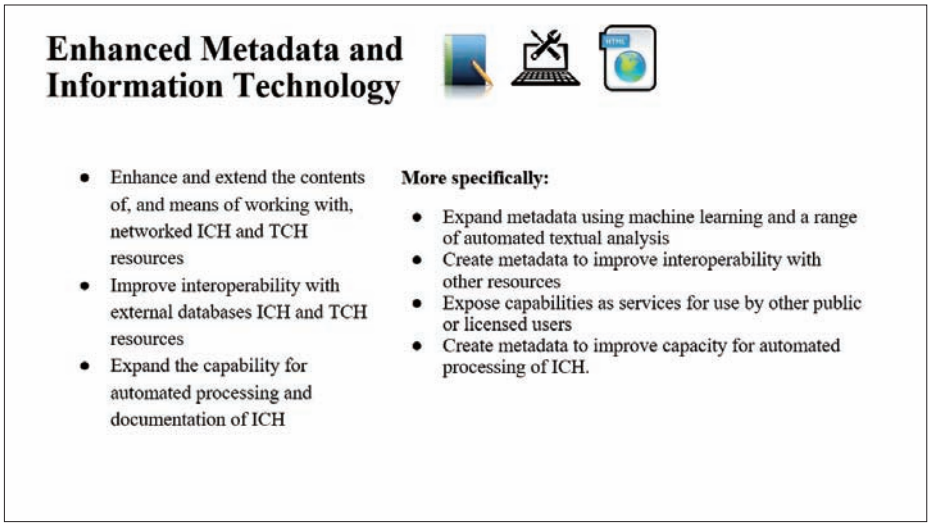


Each page contained topical metadata associated with the text, based on the Outline of Cultural Materials.



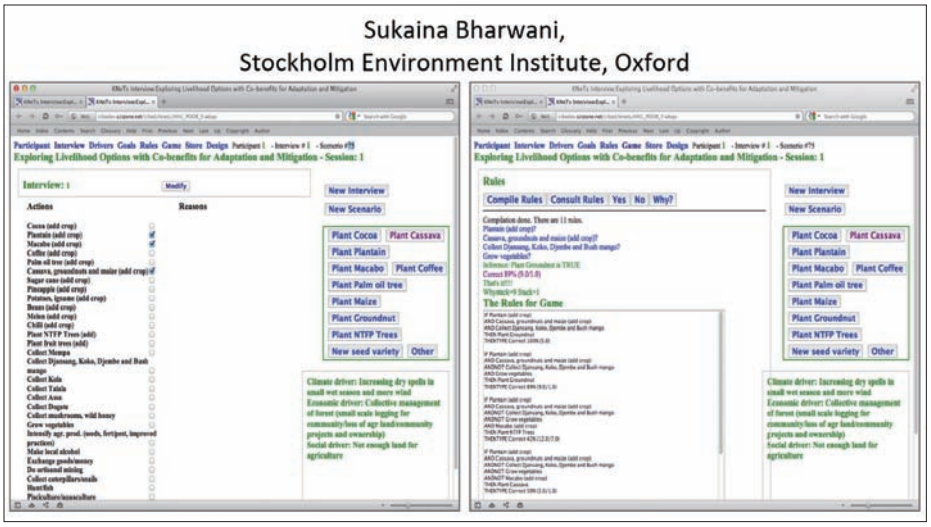
It contained other metadata across the top of the page. This insured that each page could be processed independently of its position in a particular volume.

The main reasons for declaring a "New Era of Intangible Cultural Heritage" are a greater need due to pressure on much ICH, many new forms of ICH to document, and more potential relevance for potentially any archetype of ICH to deal with future challenges we will face. In the past ICH has largely been documented and used by communities of experts, who established various roles in applying knowledge of ICH to current issues and problems. But this is an expensive and slow enterprise. For example, HRAF requires a budget of around 2 million US$ per annum to curate and disseminate its collection. The cost of collecting and producing the ICH HRAF curates, in today's dollars is in the neighbourhood of a billion dollars. And employing the current model of experts, is gravely under-utilised for a resource that could illuminate many other collections of both Tangible and Intangible Cultural Heritage.

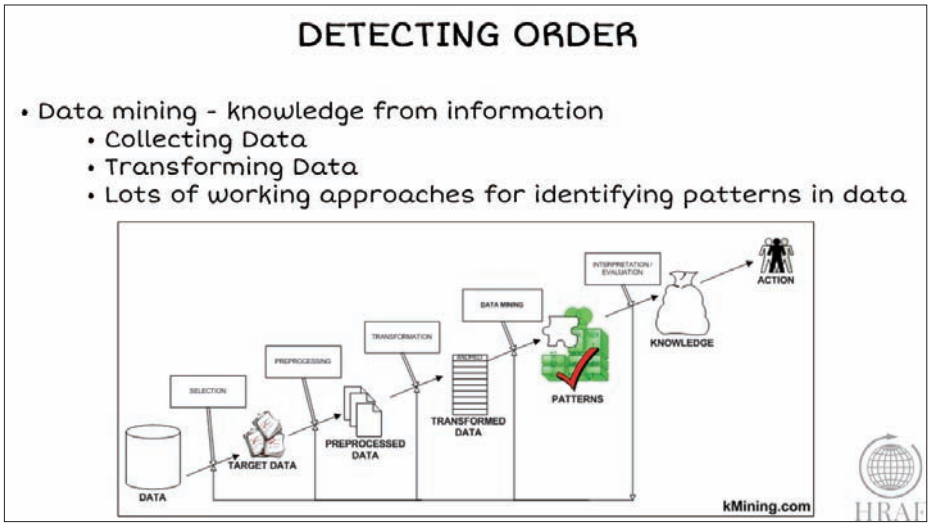


Metadata is part of the solution to breaking away from primarily direct expert mediation of ICH collection, processing and deployment, to an approach where all of these functions can be performed by non-experts. This metadata is derived from existing experts, for at least nominally metadata can represent the kinds of knowledge that experts possess about ICH. Just the development of solid metadata models for different ICH domains make the domains more accessible to non-experts.

But over the past two decades there has been a revolution in the development of machine learning for classificatory metadata and computational linguistics for treating embedded language artefacts as metadata. Perhaps, more critically, these computational tools can perform much better with the inclusion of expert-derived metadata, and with sufficient examples computational models can be evolved that can make similar judgements of similar quality to new data. At least in this small degree individual expert performances within a domain can be imparted to a computational model that can be applied to much more material than the expert could every process, and effectively put this expertise into the hands of non-experts to deploy for their own purposes.



Indeed, some of the most exciting prospects for enhancing ICH arise from the prospect of directing interacting with performers of ICH to create ICH data and metadata directly, rather than mediating between a performer and an expert. For example, Sukaina Bharwani, of the Stockholm Environment Institute, Oxford, has since 1998 been working directly with farmers and other local resource users to produce Expert System models of their knowledge relating to domains such as agriculture, water regulation and conservation or use of forest resources. This is done with a set of interviews, to collect initial judgements, and then iteratively on the resulting 'rules' that reflect local knowledge. The outcome is good, usually agreeing with the original source over 90% of the time, using rules that are unsequenced but which interact together to produce the outcomes.



Her results are achieved using a combination of machine learning to induce initial rules, and then iteratively working with the source to refine these rules for more general use. These machine learning tools work through relatively simple principles of classification, finding terms common enough to be useful, but not so common that excessive 'lumping' of cases occurs.

Computational simulations(Virtual Reality and Augmented Reality) can serve as a means of dissemination of ICH, assisting in its conservation in people, and not just in institutions. The Internet of Things is already in play, starting with phones that answer questions for you or turn on your lights. The environment will increasingly be littered with sensors and transducers to both collect information and to apply information to the environment. Machine learning and the derived metadata from this process will greatly influence both how useful (or unuseful) the IoT will be.