Research Objects: Knowledge Management for Federated Public Health Research
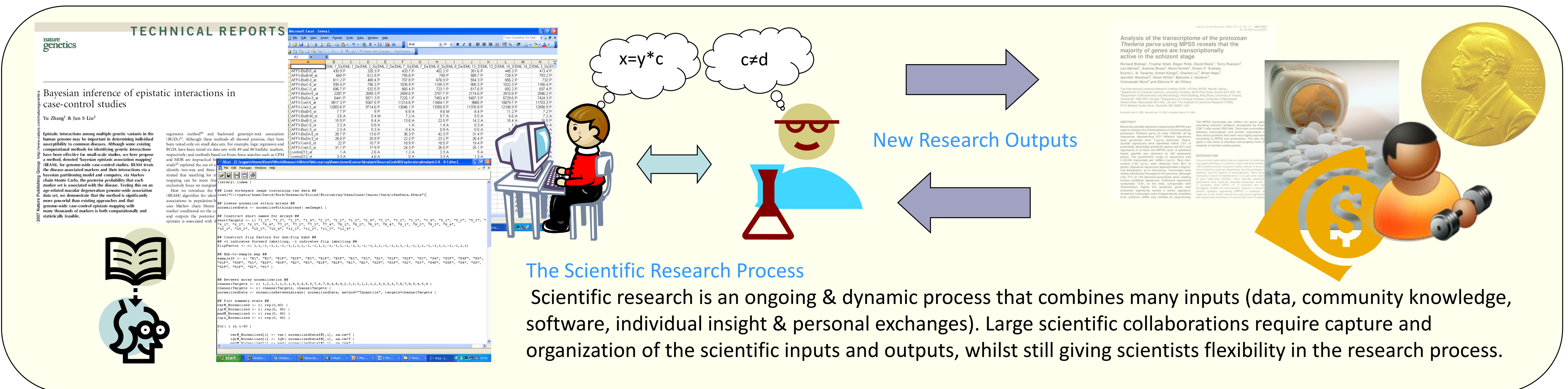
Research Objects: Knowledge Management for Federated Public Health Research

[2]Jitenkumar Bhagat, [1]John Ainsworth, [2]Carole Goble, [1]David Hoyle, [1]Iain Buchan (buchan@manchester.ac.uk)

[1]Northwest Institute for BioHealth Informatics & [2]School of Computer Science, University of Manchester

Effective management of knowledge is key to the success of projects within any research field, but especially across the broad spectrum of social to biomedical research that is relevant to public health. From the scientific experimentation phase, right up to publishing, knowledge is produced, transferred, shared, manipulated, and exposed. We propose a new model to support this lifecycle – **Research Objects (ROs),** to address the issues of manageability, reproducibility, ownership, exposure, and sharing. This model focuses on exposing research outputs and managing data assets.

ROs are containers that encapsulate the outputs of a project together with all the associated files, data objects, properties, metadata, annotations, behaviour, events, relationships, workflows and protocols required to derive the outputs. **Assets** encapsulate specific files and data objects in ROs and have associated metadata that describe their role, type and properties. They can be extended with custom actions, visualizers and processes. Whilst ROs are self-contained they allow data from external sources to be linked in, maintaining original ownership and authority.

The Scientific Research Process

Scientific research is an ongoing & dynamic process that combines many inputs (data, community knowledge, software, individual insight & personal exchanges). Large scientific collaborations require capture and organization of the scientific inputs and outputs, whilst still giving scientists flexibility in the research process.

New Research Outputs

Research Objects – A new model for managing the research lifecycle

A model for supporting scientists in working within *research contexts* and explicitly defining research outputs. A RO is an organically evolving and emerging container to capture research. In doing so we:
- Aid the scientific process by organising inputs and outputs in a flexible and extensible fashion.
- Aid collaboration through consistency in managing and exposing inputs and outputs to many researchers whilst allowing for easy capture of data.
- Enhance the preservation and reproducibility of research outputs by making the route to new findings transparent and persistent.
- Allow for tracking of the provenance of scientific outputs.
- Make the context and process of research more explicit, thereby enabling research to be conducted in sensitive data environments, e.g. Healthcare, more easily, and with greater ethical transparency.
- To replay, share, disseminate research findings in ways that allows others to customize them.
- Include social collaboration and project management techniques.

The RO model, shown right, is the conceptual model we have developed. Central to ROs is the concept of an "Asset" – a consistent manner in which to handle different forms of data.
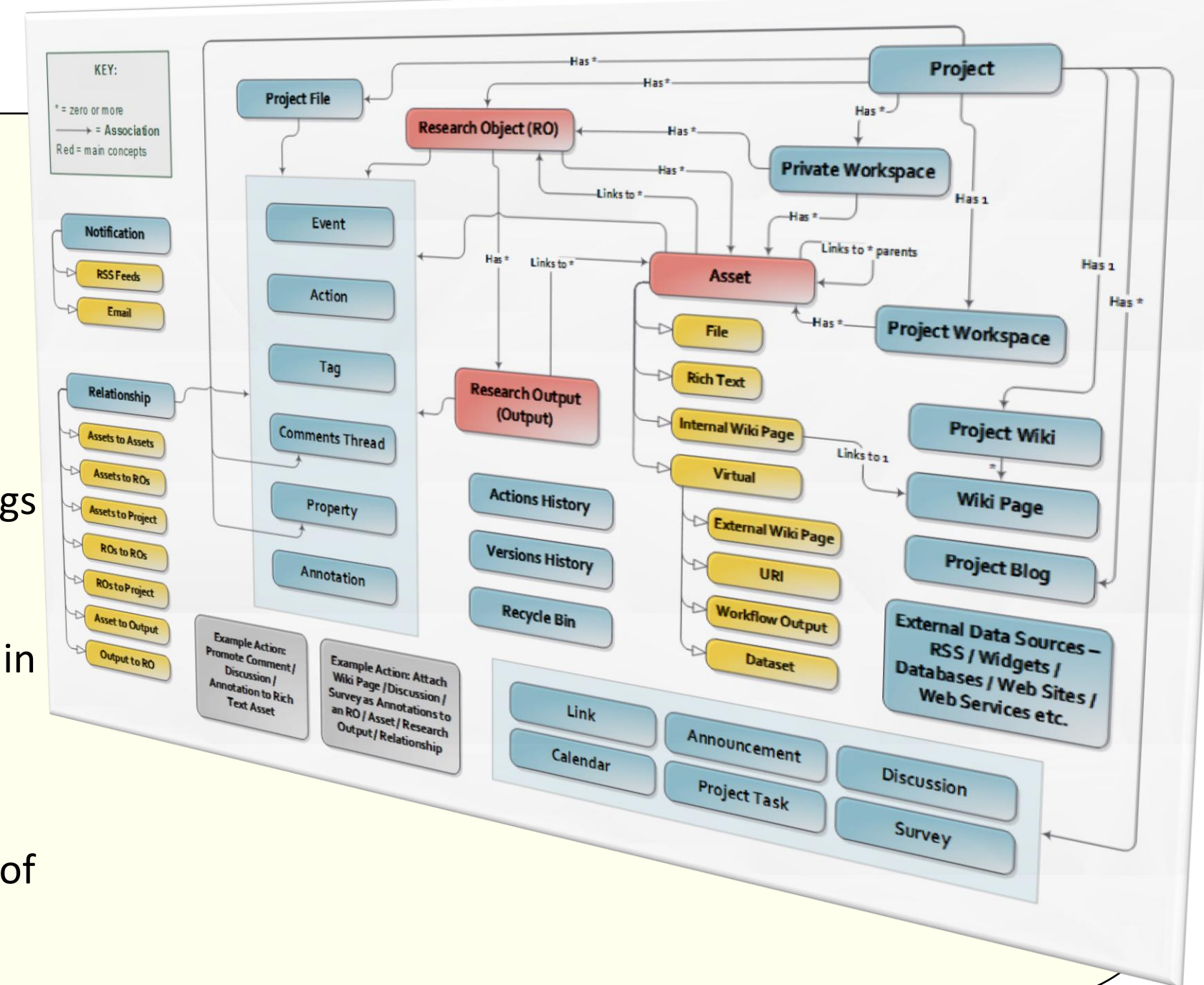

Working with Research Objects

We need some method of manipulating and exposing ROs to researchers.
- Collaborative technologies, e.g. Microsoft SharePoint™, are often primarily influenced by the need for developing enterprise solutions.
- Enterprise orientated collaborations often have well defined goals and processes.

Scientific research collaborations can be much more dynamic, with only vaguely specified end-points and methodologies. Therefore the RO conceptual model acts as a *buffer* and offers concepts that are more suited towards scientific research. We have developed a 'dashboard' user interface for managing projects using ROs, on top of the core SharePoint™ functionality.

The object-centric approach allows social collaboration features to be added. For example, tagging and commenting of Assets and ROs can enhance the way people collaborate and promote better communications between researchers.
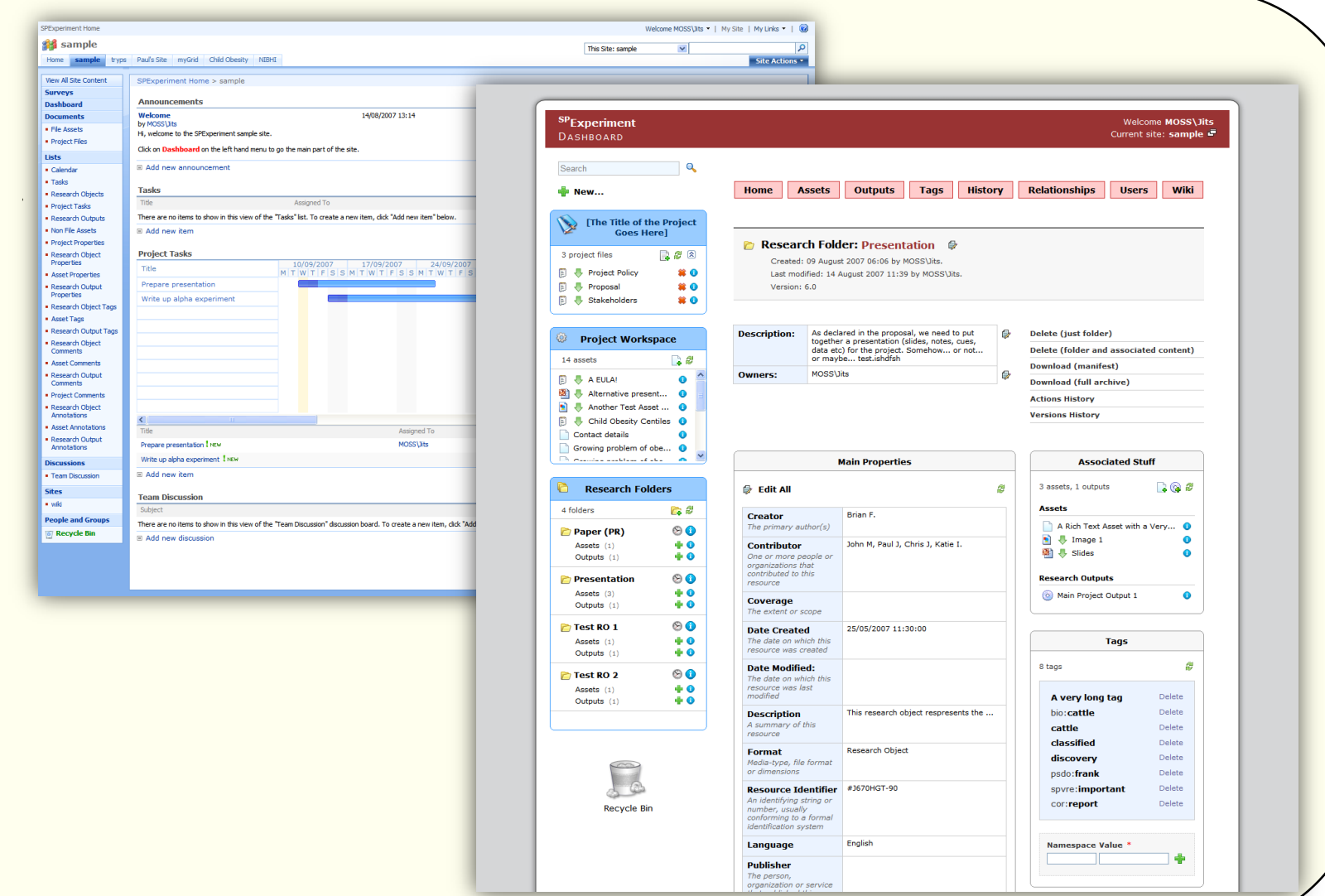
Future

The full benefits of the Research Object (RO) paradigm will only be realised when ROs can be generated, transferred and interpreted by the systems used to support each stage of the experimental process. To this end the systems that comprise an e-Lab will use Research Objects as the lingua franca. We also envisage Active Research Objects where a Research Object is combined with Agent technology, to create a semi-autonomous entity capable of interacting with and travelling between systems.