

CERIF, CRISes and Research Databases

Dr Mark C. Cox
King's College London



King's “CRIS”

- Database holding various types of research-related information – the “Research Gateway”
- Developed 2004-7 (although development continues as part of VRE project)
- Oracle DB (reflects other business systems)
- Web (php) interfaces for administrative and academic staff
- Devised as general research data management system, but focussed on RAE

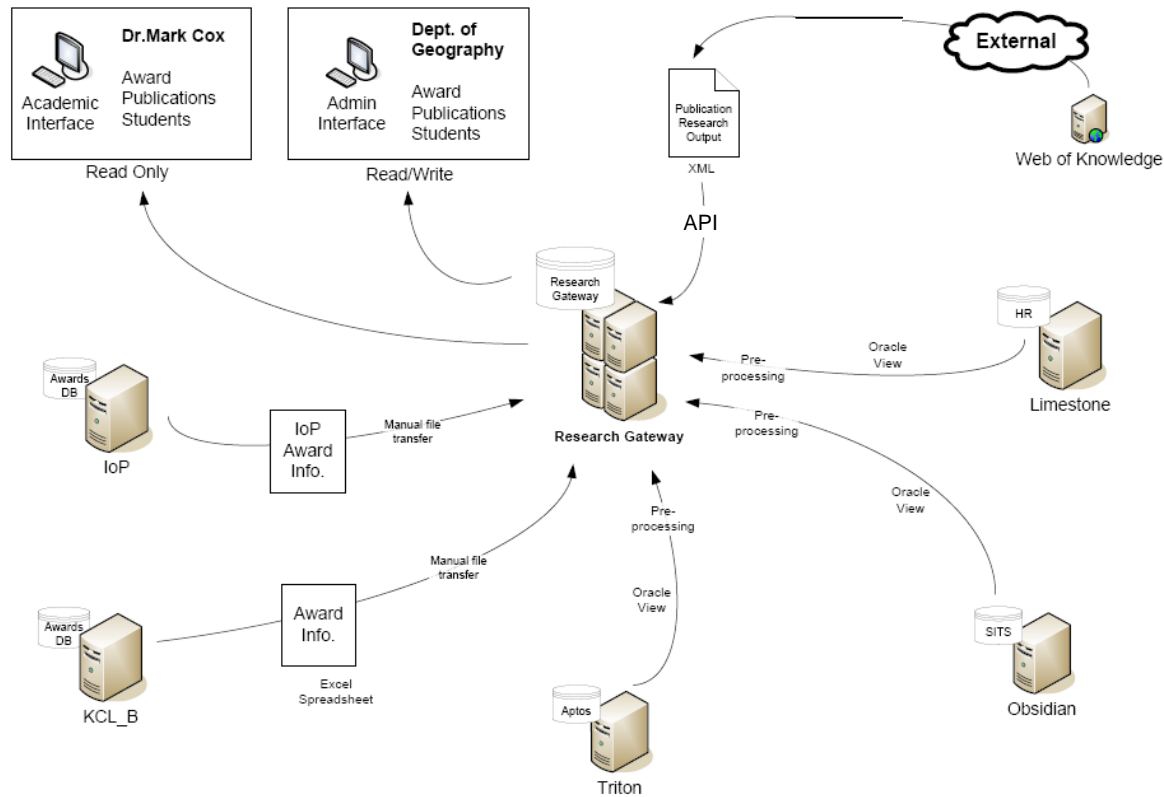


King's Research Gateway

- Data feeds from:
 - HR system
 - Student system
 - Awards databases
 - Finance system
- Also publications feed from Thomson Reuters' Web of Knowledge (WoK)
- Workflow to ensure that data changes are made in "primary" system
- Avoids need for multiple changes to same data



Research Gateway model





CERIF

- Common European Research Information Format
- International-standard data model which describes Research Entities and their relationships
- Developed originally with the support of the European Commission
- Development now under the custody of euroCRIS
- CERIF is neutral to architecture



Common European Research Information Format



© Brigitte Jörg



Readiness for REF (R4R) project

- One strand of the project is the development of a bespoke version of CERIF - CERIF4REF
- This will encompass the types of data that will be needed for REF submissions



CERIF4REF

- REF data requirements not yet finalised
- Documentation suggests considerable overlap with RAE requirements
- Initial mapping exercise was carried out to map RAE 2008 data fields to CERIF elements
- Data fields taken from RAE forms
 - RA1: Person data
 - RA2: Research Outputs
 - RA3a/b: Students and Studentships
 - RA4: Research Awards



CERIF4REF

- First mapping exercise was reviewed and refined by euroCRIS developers
- Vast majority of RAE data fields could be mapped to CERIF elements
- Some missing data (DoB, FTE) are likely to be incorporated into next version of CERIF
- Summary fields needed to be re-expanded to basic elements
- Thus RAE data fields required aggregation of these CERIF elements

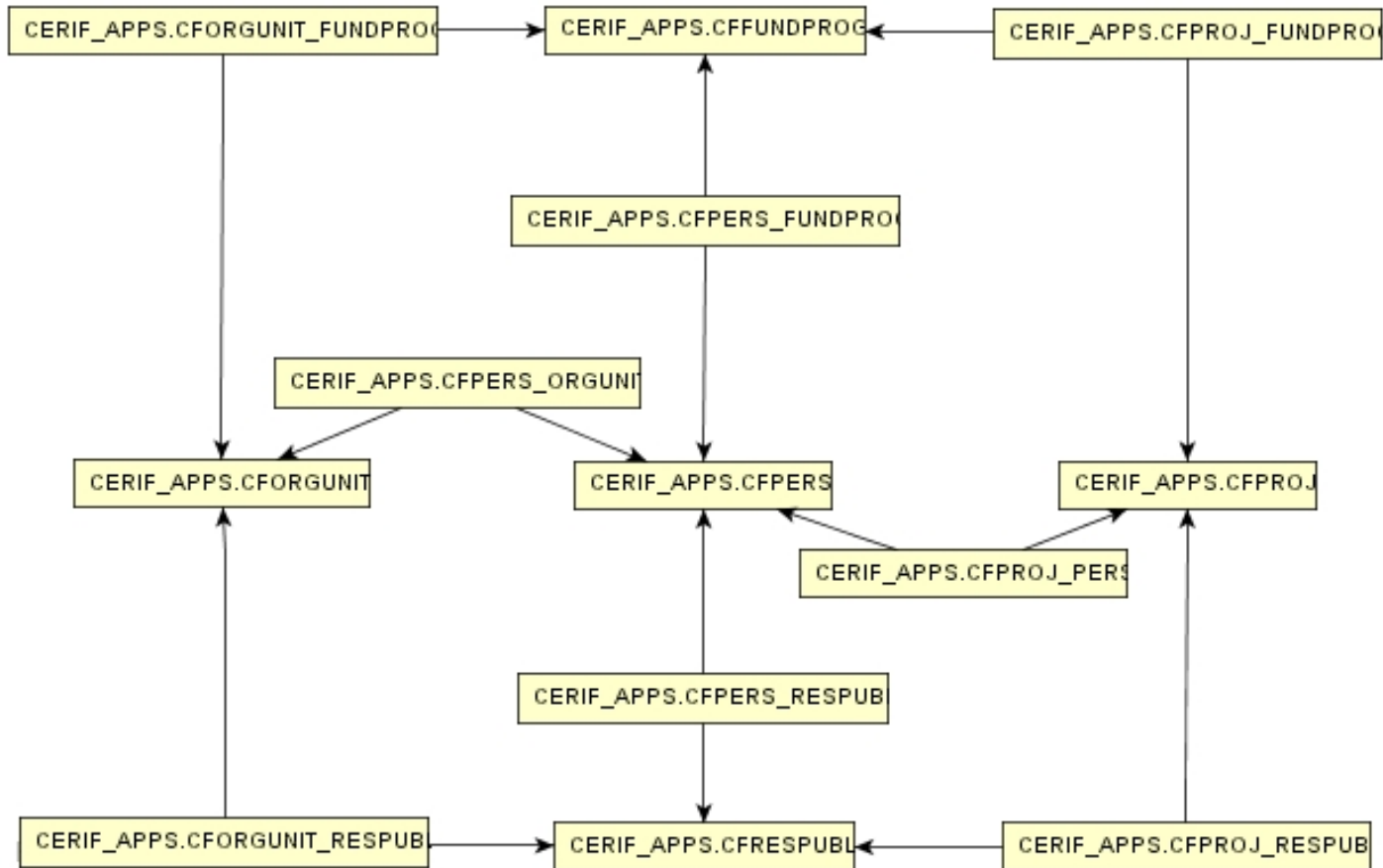


Creating a CERIF CRIS for King's

- To allow CERIF benefits to be realised, our aim is to transform RG data into a CERIF compliant format
- The first stage was to construct a new Oracle database conforming to the CERIF data model
- This was done by using SQL scripts created by euroCRIS
- Scripts have been written for many common databases



Oracle CERIF – core tables





Populating Oracle CERIF DB

- Initial attempts to populate the Oracle-CERIF database were made via the RAE forms (RA1, RA2, etc)
- CERIF4REF mapping was used to match RAE form fields to CERIF elements





Populating Oracle CERIF DB

- RA1 data: relatively straightforward, with most elements mapping to CERIF; DoB and FTE, however, do not yet exist in CERIF schema
- RA2 data: more difficult to map – RA2 fields contain different data elements depending on output type
- RA3a,b, RA4 data: in these forms, data is summarised – without the underlying data the relevant entities cannot be mapped

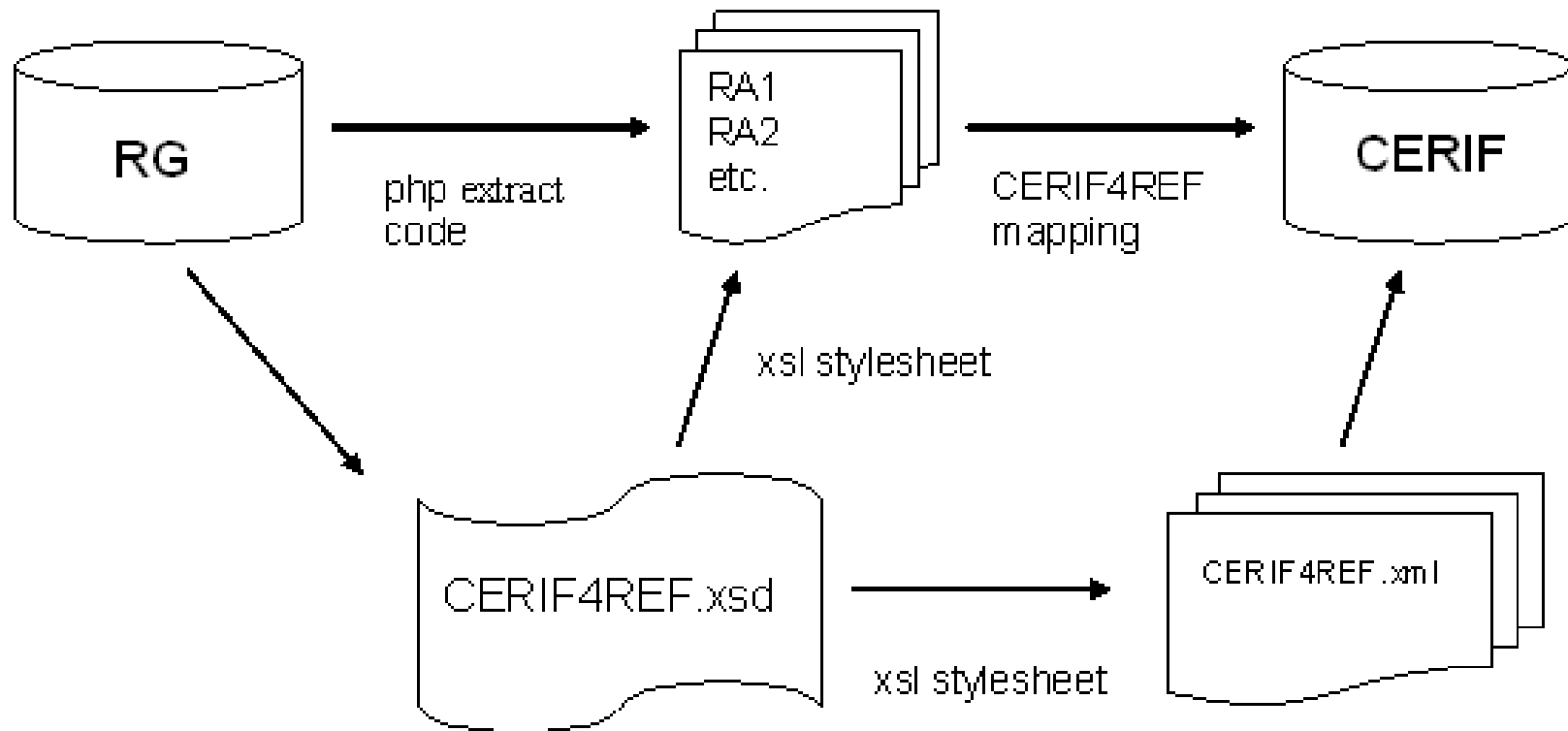


CERIF4REF schema

- To aid with defining elements, an XML schema (.xsd) has been created for CERIF4REF
- This describes the underlying elements that are required to generate RAE/REF summary data, such as RA3 and RA4 forms
- Stylesheets (.xsl) have also been generated which can be used to transform CERIF4REF data into pure CERIF format, or into RAE forms.



Populating Oracle CERIF using schema





Using CERIF4REF schema

- Benefits
 - All underlying entities can be defined, e.g. individual students and awards
 - Once mapping is complete, RAE forms generated from stylesheets can be compared with those generated by php scripts to check accuracy
- Drawbacks
 - A separate mapping exercise is required to map RG data onto CERIF4REF schema; however, elements are much more similar to those already defined in RG



Next steps

- Once mapping is complete, population of Oracle CERIF database directly from RG
- Generation of real King's data in CERIF xml format
- Exchange of data with other CERIF compliant systems