Project Agreement Number: 519138-LLP-1-2011-1-UK-KA3-KA3MP
Project funded by the European Commission

Deliverable D2.1 Edition 2

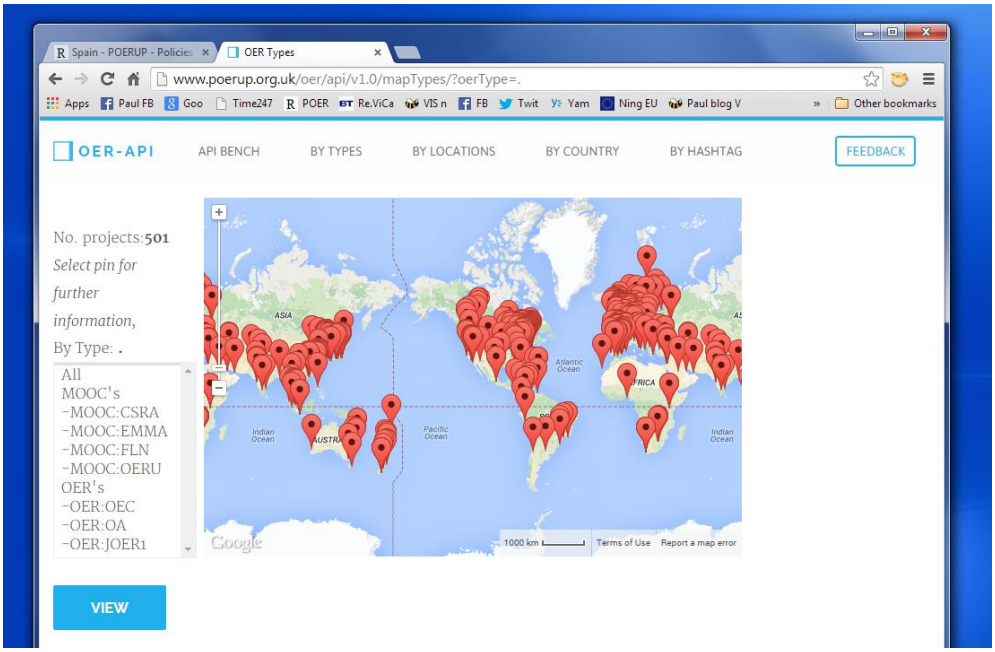
Transversal and categorised inventory of OER Programmes and Initiatives – on maps

Document information	
Due date of deliverable	original edition delivered on 31 July 2013
Actual submission date (final version)	this revision delivered on 30 June 2014
Organisation name of lead contractor for this deliverable	Sero Consulting
Revision	Edition 2

Dissemination Level		
PU	Public	☒
PP	Restricted to other programme participants (including the Commission Services)	☐
RE	Restricted to a group specified by the consortium (including the Commission Services)	☐
CO	Confidential, only for members of the consortium (including the Commission Services)	☐

Contents

The Sero Custom OER Map



Executive Summary

POERUP is an EU-funded project whose funded period ran from November 2011 to June 2014 inclusive. Its purpose was to develop OER-friendly policy recommendations, based on analysis of existing OER initiatives, countries, policies and case studies. For POERUP Work Package 2, Sero has created a curated map/database of 501 open education initiatives, both OER and MOOC.[1]

A companion but much smaller work item created a report on OER *policies* – for details see *Deliverable 4.1: Overview of European and International policies relevant for the uptake of OER*.

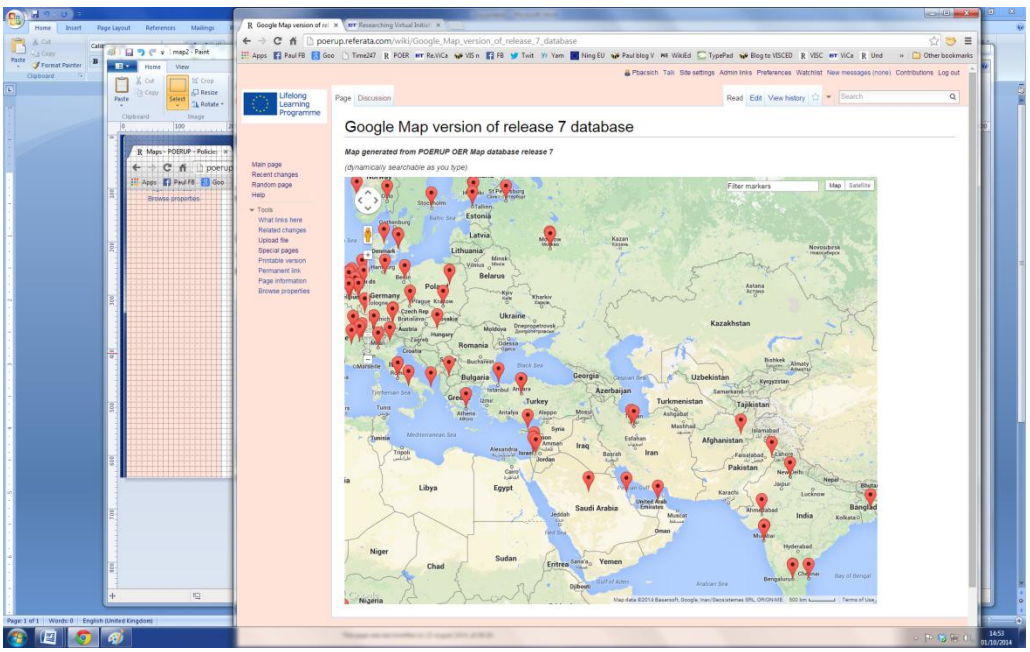
First, Sero created a *Custom Map Tool* driven by the sophisticated "noSQL" database MongoDB to allow display of and search for OER initiatives – http://oer.poerup.org.uk – as part of a wider initiative to document and allow search for open education initiatives, including MOOCs, available at http://www.poerup.org.uk. The core database technology and approach were chosen to be scalable to high performance as well as being open source and Linked Data-ready. The Open API it makes available facilitates future use by different groups working collaboratively on problems of collecting, mapping and analysing open education initiatives, including but not only eMundus, SharedOER, D-TRANSFORM, OER Africa and, possible Hewlett-funded initiatives in this area.

Second, Sero created *Semantic Map Tools* using Semantic Maps, a module of Semantic MediaWiki, hosted on Referata to support the POERUP wiki – http://poerup.referata.com. Semantic Wiki is a powerful extension of the MediaWiki software. (MediaWiki is used also for WikiEducator and Wikipedia.) See for example the map at http://poerup.referata.com/wiki/Languages

Third, Sero created a number of *Google Map Tools* consisting of ad hoc maps and charts, using Google Map Engine Pro and Google Charts – e.g. the "POERUP 501" map at https://mapsengine.google.com/map/u/1/edit?hl=en&authuser=1&mid=zYG2prGO09jE.kj-xjWvhQNjg

This deliverable was written at the very end of the POERUP project and so inevitably has aspects of looking to activities beyond the project and in particular Exploitation of project results.

A Sero Google Map loaded from the Excel OER database



[1] This will in Autumn 2014 be extended by at least 100 more initiatives, in the exploitation phase of POERUP.

1 Introduction and the aims and objectives of this document

This is the second edition of Deliverable 2.1 of Work Package 2 of POERUP. The Deliverable Title from the proposal is:

> *Transversal and categorised Inventory of OER Programmes and Initiatives*

The Work Package title is:

> *Cross-sector analysis and comparison*

Extracts from the brief for the Deliverable observe:

> *In the first phase the research team will build upon the findings and results of relevant (inter)national/regional/major OER initiatives. The desk research will include a meta-analysis of previous publications, research and activities in the field and take into account previous project results concerning OER. We have already identified more than sixty sources both from our own work (http://poerup.referata.com) and from existing projects including OPAL, OLCOS, EDRENE and OER HE. We will also draw on the knowledge gained in the Re.ViCa project (http://revica.europace.org) (on virtual campuses in universities) and ongoing in the VISCED project (http://visced.referata.com, virtual schools and colleges) in creating a cross-sector categorised inventory of OER initiatives and programmes in Europe and the rest of the world. The policy advice will provide them with an in-depth understanding as to the importance of, amongst other factors, the policy context. In particular, an analysis of past policy-relevant successes (and any failures we can discover) will make a significant contribution towards better decision-making by this target group.*

And later:

> *With the cross-sector inventory we want to go further than the traditional list of good practices so as to combine all relevant parameters: financing, educational structures, pedagogical approaches, quality procedures, content production methods, business models, organizational embedding, the role of communities and sustainability, etc. Based on the results a categorisation will be made. This will include a classification by the political scale of initiatives: National initiative (e.g. Wikiwijs) – but also regional and international. Partners have much experience on this from the prior/ongoing projects Re.ViCa and VISCED as well as national work (e.g. for Becta).*

A useful and influential first edition of Deliverable 2.1 was produced in July 2013 by Ming Nie, then of the University of Leicester, with the advice of a team from Sero. But as noted in the deliverable template, it was expected that this would be "updated through the entire project". As the second round of country report updates got under way in early 2014, it became clear not only that there were many relevant initiatives to add, but also that the original printed tabular Word representation of initiatives was no longer feasible. As there was much influential thinking going on in other parts of the world to put OER initiatives on to geographic maps (see in particular http://unescochair.athabascau.ca/oer-mapping-exercise), POERUP decided that a mapping focus was the best approach.

*Note that this deliverable is formally a **second edition** of Deliverable 2.1. As noted in the original POERUP work plan: "The aim is to have at least 100 major initiatives collected, documented and classified." We have considerably surpassed that threshold.*

2 Background on Sero and those involved in the map

Since this document is designed also to be the basis of more than one self-contained report, we have included some additional information in it not normally found in deliverables.

2.1 Sero Consulting Ltd

Sero is an education consultancy specializing in assisting institutions and government agencies with the exploitation of IT including learning resources (libraries) and e-learning. Formed in 2004, Sero has 20 staff and associates, including Dr Paul Bacsich (work item leader), and Dick Moore of Moore Answers. Sero has many years' experience in working with non-profit organisations and foundations – including the European Commission (and its agencies and research labs), national, regional and local governments, JISC, and the UK Higher Education Academy.

Sero is the Project Manager for the EU project POERUP – Policies for OER Uptake – which has provided funding for this work item – and was the Project Manager for the project VISCED – Virtual Schools and College Education – http://www.virtualschoolsandcolleges.info. Both these projects developed large wiki databases of educational initiatives at the institutional and programme level. They also leveraged on the prior project Re.ViCa (2007-2009), in which Paul Bacsich played a key role, which created a large wiki of virtual campus initiatives – http://www.virtualcampuses.eu. Many of the issues of how to describe, name and classify countries, regions, institutions and initiatives are solved – this work was done first in Re.ViCa (2007-2009). In VISCED and POERUP Sero had to handle a worldwide network of consultants, partners and volunteers making input – see the *People of POERUP* map – http://bit.ly/peopleofpoerup.

2.2 Staff and roles in WP2 work item on OER Map

The key staff for this WP2 work item were:

1. Work Item Lead and database design: Dr Paul Bacsich, Senior Consultant, Sero Consulting Ltd, Project Manager of POERUP and designer of the Sero OER Map Tools.
2. Information Architect: Dick Moore, formerly Director of Technology at Ufi/learndirect (a leading online provider) and the implementer of the Sero OER Custom Map tool.

Dr Paul Bacsich

Dr Paul Bacsich is Senior Consultant at Sero. He was at the UK Open University for 24 years, ending as co-founder/Assistant Director of the Knowledge Media Institute. He then became full Professor at Sheffield Hallam University where he set up an early Virtual Campus. As well as his research in education, he is an expert on the web and data communications, with an increasing focus on web science. His mathematical background of a PhD and post-doctoral research in mathematical logic has been helpful for his work on semantic web and Semantic Wikis.

Paul started comparative research on online learning in 1995 – http://www.pjb.co.uk/6/visit.htm. During 2007-2009 he played a key role in the EU project Re.ViCa – Review of Virtual Campuses – which developed a large wiki database of initiatives in post-secondary online education. In 2010-13 he led the EU project VISCED (Virtual Schools and Colleges) which reported in February 2013. He then bid for and now leads the EU project POERUP which is due to report at the end of September 2014. His role is as Project Manager and leading the collection of OER initiatives around the world, as reported on in detail in the POERUP wiki at http://poerup.referata.com/wiki/Country_reports.

In August 2012 he spent six weeks in New Zealand as the guest of one of the university partners of OER u. In August 2013 he co-led an IT in education workshop in Brazil, with Casey Green and Abdul

Khan, for the leaders of private providers of HE. He has much other international experience including many visits to Canada and the US. (He spoke at iNACOL in 2012 and has attended several Educause meetings.) He has personally geocoded all the open education initiatives in the Sero OER Map database and generated the region information, giving him great insight into these aspects.

His roles in this work item were: Item Lead; designer of the database structure for OER initiatives; developer of the maps using Semantic Maps and Google Map Engine Pro.

Dick Moore

Sero Associate Dick Moore has 12 years experience at board and director level in Ufi/Learndirect, a major online provider of vocational education. He is now runs a consultancy *Moore Answers* for a wide range of organisations within the educational, business and not-for-profit sectors. He has practical experience in building, managing and delivering enterprise-class, service-focused, innovative internet solutions. Headhunted as Vice-President Systems and Service at "the Dock", a dot.com based in Los Angeles, he created a technical platform and service team from scratch in under 3 months. As Director of Technology at **learndirect** he led the transformation and re-engineering of the service from one that was poorly performing and costly into one that remains a world-class, leading-edge service that was a finalist in the Orange Business Awards and won the UK's E-Government *National Award for Innovation* in 2009. As a Trustee of the Association for Learning Technology (ALT) he oversaw the transition of the Association's journal ALT-J to open access. He is also on the Technology Advisory board for the Open College of the Arts. He has a particularly topical interest in extracting usable data from Wikipedia and other wikis. He was the developer of the OER Custom Map Tool for POERUP.

His role in this work item was Information Architect and lead developer of the Custom Map Tool.

Other staff

Valuable advice was provided by Dan Wilton of Athabasca University, based on his experience with developing the eMundus Atlas – http://emundusatlas.org.

3. Focus area for data collection

In the POERUP project, OER and related initiatives (such as MOOCs) were collected from *all* sectors of education and from *all* countries. There was an *implicit* focus towards the more pedagogic uses of OER, with less focus on collection of institutional and publisher open access repositories of mainly research interest: these are already collected and curated by the OpenDOAR initiative based at the University of Nottingham, UK – http://www.opendoar.org – and prototype harvesting of these has been done by Sero – see http://bit.ly/opendoarafricamapbysero.

The POERUP country studies provide global coverage of OER-related initiatives until July 2014 across all education sectors. The 33 countries specifically covered by POERUP are:

1. EU/EEA (17 out of 31): Belgium, Denmark, Finland, France, Germany, Greece, Hungary , Ireland, Italy, Netherlands, Norway , Poland, Romania, Portugal, Spain, Sweden, and UK)
2. The Americas: United States, Canada, Mexico and Argentina
3. Australasia: Australia and New Zealand
4. Middle East: Saudi Arabia, Gulf States (5) and Jordan
5. East Asia: Thailand
6. Africa: South Africa and Rwanda

In addition, POERUP analysed countries studied by UNESCO IIEP (France, Lithuania, Russia, China and Brazil) and OER Asia, as well as some other reports (such as Turkey). Four sub-continental sweeps

were also done: Hispanic America; Commonwealth Africa; Asia; and Yugosphere (countries of the former Yugoslavia). Specific features of the database include full coverage of all OCW institutions.

Over half of these country studies were done by staff at or consultants contracted by Sero.

4. Methodology

This WP2 work item work plan comprised the following six tasks across two phases:

1. POERUP phase: Sub-project set-up
2. POERUP phase: Identify key technologies – for databases, maps etc
3. POERUP phase: Database design
4. Post-POERUP phase: Database design update
5. Post-POERUP phase: Rendering updates
6. Post-POERUP phase: Editor interface

4.1 Sub-project set-up

The work item was commissioned in January 2014, with a burst of development activity in May-June 2014.

4.2 Identify key technologies

It was decided early on that the Sero Custom Map Tool would use as its database system the open source *noSQL* database MongoDB. Mongo has substantial advantages of flexibility and performance for retrieval situations and can scale indefinitely. It was chosen in favour of a specific SQL solution because of the need for flexibility in database design: among its advantages are that it can handle multiple record types (thus slightly different record types for initiatives, policies, institutions and individuals) and fields with multiple items (array fields). It is hosted in the cloud – initially this is on Amazon Web Services.

As noted on http://www.mongodb.com/mongodb-overview:

> *MongoDB is an open-source database used by companies of all sizes, across all industries and for a wide variety of applications. It is an agile database that allows schemas to change quickly as applications evolve, while still providing the functionality developers expect from traditional databases, such as secondary indexes, a full query language and strict consistency.*
>
> *MongoDB is built for scalability, performance and high availability, scaling from single server deployments to large, complex multi-site architectures. By leveraging in-memory computing, MongoDB provides high performance for both reads and writes.*

It allows fast search on any field using secondary indexes. It also has features such as auto replication; sharding; arithmetic, string and geospatial functions; and JSON document-focused structure – http://www.json.org.

The team provided a Restful API – http://www.restapitutorial.com – to enable multiple different user interfaces to be developed.

Beyond the end of the POERUP project, the POERUP wiki will be used longer term for data collection as the focus inevitably moves from "sprint" collection of groups of initiatives by specialists to ongoing collection from a wider open education community. This necessitated the development of a

wider set of tools more integrated with the wiki. The wiki also supports harvesting of information in JSON, XML, CSV and a range of other formats.

4.3 Database design

The database design went through several iterations. It is now a slight update of the "production" database design used in the Custom Map Tool – http://oer.poerup.org.uk – and the Excel database that "feeds" it. The following describes the production database rendering in MongoDB (as seen on the Custom Map Tool):

1. Accession number (accno): numeric (8 digits, no leading zeros)
2. Hashtag: a series of letters and digits – e.g. *futurelearn*
3. Type: followed by subtype – e.g. *OER:OEC* for a member of the Open Education Consortium
4. Countries: the country of the headquarters – as in ISO 3166-1 alpha-2
5. Region: the region where the headquarters is – as in ISO 3166-2
6. City: city/town where the headquarters is based, using the name in English as in Wikipedia
7. Name: name of the initiative
8. URL: one only
9. Summary: a *one-paragraph* abstract of the initiative
10. Owner: the institution who owns the initiative
11. Address: the postal address of the owner, in standard form focussed on a physical address
12. Geocode: latitude then longitude in decimal degrees, separated by comma as standard
13. Contact: in *<firstname> <surname>* format
14. Email: email address of the contact
15. Scale: in a one-word vocabulary
16. Funders
17. Start year (start)
18. End year (end): if it has ended
19. Levels: controlled vocabulary – HE, FE (VET), schools, adult, etc
20. ISCED: numeric values taken from ISCED 2011 (replacing ISCED 1997)
21. Interface language (intlang): the language that the interface is in
22. Resource language (reslang): the language(s) that the resources are in
23. Subjects: free vocabulary, items separated by semicolons
24. License: Abbreviated form, e.g. CC BY-NC-ND
25. Media: types of media, e.g. Text, Video
26. Tags: at the curator's discretion
27. Latitude (lat): calculated from geocode
28. Longitude (long): calculated from geocode
29. georegion: a variant of the UN geoscheme more appropriate to OER
30. Prime: first country in the list at Countries in field 4
31. Code: ISO 3166-1 alpha-2 code for that country, e.g. AU for Australia
32. Continent: the first part of the georegion
33. Zone: the second part of the georegion
34. Canonical: the ISO 3166-1 alpha-2 name of the first country entered in field 4
35. A number of internal fields to do with database audit trails and display aspects.

The API allows you to extract in JSON format individual and sets of records for consumption within third party tools including Excel or open source spreadsheets – or indeed into Google Map Engine Pro: see the copy of key fields from the released database at http://bit.ly/poerup501.

The collection version of the database (currently in Excel) has shortly before the end of the project been upgraded to store more detailed regional and language information as follows:

1. rcode – the ISO 3166-2 code for the region
2. canreg – the canonical name, usually in the local language if that uses roman script

The Region code (rcode) is based on the standard list of regions and their codes at ISO 3166-2, thus *US-CA* for California; with canonical name for the code (as given in ISO 3166-2). The regional coding will considerably facilitate and standardise future "search by region" features, when such search functions are added.

The original language field has recently been joined by four new language fields:

1. lcode – Language code as per ISO 639-1, such as *ar* for Arabic or *lv* for Latvian
2. alpha3 – Language code of three letters as per ISO 639-3 such as *afb* for Gulf Arabic or *ltg* for Latgalian (within the Latvian macrolanguage) – this is required for the kind of discrimination that projects such as LangOER wish to make (http://langoer.eun.org)
3. canlang – Canonical name for ISO 639-1 2-letter code
4. canlang3 – Canonical name(s) for ISO 639-2/T alpha-3 3-letter code

This will considerably facilitate and standardise the "search by language" feature when implemented. Note that the two-letter codes are the ones used in Wikipedia language versions and are much more familiar to users than the three-letter codes.

4.4 Database design updates considered

Subregions and alternate regions

The sub-project considered the usefulness of a subregion field **sregcode** – Subregion code – as given in ISO 3166-2. An example is Spain: *ES-T* for Taragona (province) within *ES-CT* (Catalunya). Given the complexities of this aspect and some cost issues (there appears to be no open source version of the standard usable by developers: it has to be purchased from ISO), it was decided to defer implementation until after the end of POERUP.

Consideration of subregions brings one of the issues which do not become evident until one has geocoded many initiatives. Note that unlike the stability in state/provincial boundaries in the US, Canada or Australia, there is much less stability in regional names or boundaries in Europe. In several European countries there is a concept of "traditional province" (or "county") which no longer conforms to any political divisions but may have strong linguistic/dialectal or cultural significance. In other countries, as now in France, the lower levels (departments) are more stable but the top level province structure more prone to reorganisation. There are various other anomalies: in New Zealand the first level subdivision is to the two main Islands, which have no legal or political relevance. Similar situations pertain in countries as diverse as Indonesia and England. Thus consideration was also given to the concept of an "alt-region" to cope with these situations where the ISO 3166-2 classification seems unsuitable. In the EU and OECD the NUTS-1 classification is useful – http://en.wikipedia.org/wiki/Nomenclature_of_Territorial_Units_for_Statistics. Again, it was it was decided to defer implementation until after the end of POERUP.

Languages

Within the current academic year (2014-15) it is not expected that even more sophisticated discrimination will be required by the typical user; though some pressure may come later from specialised OER projects and OA repositories which have a focus on ancient languages. The *glottocodes* approach is there for use at that point – http://glottolog.org/glottolog/language – and already within a Linked Open Data paradigm. This is not expected to be a near-term requirement – Paul Bacsich is on the Advisory Group of LangOER and will monitor this situation.

Educational levels

Even the newer ISCED 2011 classification does not provide sufficient discrimination for classification at the K-12 level, where the chronological age of the student is felt by many, but not by some – those keen on competences – to be a key factor. Thus we considered adding a **K-field** which will be used only when the ISCED field is 1, 2 or 3 – with the K-field having values in the range 1-12 (age or competence level). Even though the "K-12" nomenclature is not used outside the US and a few other countries, most educators are familiar with the concept of the "grade" of school education. Again, it was it was decided to defer implementation until after the end of POERUP.

Educational subjects

There is already a **Subject** field in the production database, free text. It is easy to implement search on this field. In an ideal world, it would be useful to develop a controlled vocabulary or use one, but although such exist, we currently do not feel that the OER community is ready for this approach. When it does, the ISCED-F 2013 subject classification is waiting. If the database evolves later into a federated search engine for *content*, this issue *must* be revisited. (OpenDOAR has a simple structure of Subject Area Codes – http://www.opendoar.org/tools/api13codes.html – which may be relevant.)

Again, it was it was decided to defer implementation until after the end of POERUP, but we expect to implement this within the next 12 months from the date of this report.

Georegions

At present our division of the world into continents and georegions follows pragmatic, political and educational considerations – so, like others, does not always follow the strict UN geoscheme. It is easy to include the UN geoscheme if some users need it: all the mappings from countries into georegions are table-driven and easy to change. For example it would be easy also to add linguistic or political regions (e.g. Commonwealth, Francophonie) rather than being constrained by geography. Such features are likely to be required within the next year, in particular for SharedOER.

4.5 Rendering updates considered

Pin clustering

In situations where pins cluster closely, such as many initiatives in a small island (e.g. Singapore) or at one university (e.g. Athabasca University or UK Open University), the current Custom Map Tool does not group overlapping pins into one mega-pin (as is now done for example on the eMundus Atlas or http://oermap.org). This technology is not hard to implement. However, within the clustering option it may be felt useful that when there are several initiatives at one institution, their precise geocodes are adjusted *not* to coalesce at maximum zoom. Existing maps do not seem to do this. Of course most existing maps have too little data yet for this to have become a problem.

Pin colour and shape

More attention will be paid to user feedback on appropriate colour and shape of pins, e.g. as seen in many maps using Google Map Engine. Several of these rendering ideas, e.g. clustering and colours, are now available in the Google Map Tool and Semantic Map Tool renderings of all or parts of the database. See Sero's http://bit.ly/poerup501 map and its new map on Open Access in Africa – http://bit.ly/opendoarafricamapbysero.

Mobile-friendliness

All the Sero tools need improvement in respect of mobile-friendliness. The Custom Map Tool was developed using an adaptive framework (bootstrap) and has a mobile friendly user interface using buttons rather than links, but more is expected to be done in this area after the end of POERUP.

4.6 The Editor Interface

The Custom Map Tool contains a comment box to allow general comments on the Tool and also a "new OER" function – http://oer.poerup.org.uk/new_oer/ – asking the user to supply:

1. Name of project
2. Project URL
3. Project description
4. Institution name (that owns the project)
5. Contact name (at the project, or someone else who knows about the project)
6. Institution address (used to calculate the Map pin)
7. Project type (OER or MOOC, at this stage).

While this may seem a rather minimal set of fields, in our experience so far, once an analyst has the URL (field 2) and clues about the owner (field 4), it is usually easy to find out the rest – though for some countries and languages it is still hard to extract address information from institutional web sites. Several of the other fields (not only geocoding) are best left to professional curators.

Having said that, future curators will certainly want screen editor access to all (user-enterable) fields in the database; and in particular they will want to update geocoding (some gets done hastily, maybe just coding the city) and also maybe to nudge geocoding to avoid clashes. It is likely that an existing open source web-based editor such as Mongo Edit will be deployed – little implementation is required.

However, there is also an increasing need in future projects to have a work stream to cater for the following situations:

1. corrections or update to *existing* entries, such as a change of contact name (quite frequent) or a change of departmental address
2. catering for less sophisticated editors than curators will be
3. supporting a move to a more sustainable regime of a constant flow of small updates to entries rather than the current "collection sprints" typical of POERUP and its consultant teams.

Thus Sero carried out some development work to implement a forms-based editing interface suitable for a wide range of users that also *tracks* edits and updates to entries. The natural way of doing this, and in our view the most consistent with the MongoDB key/value approach, is to use semantic wiki technology, in particular Semantic MediaWiki.

Sero was familiar with the Semantic Wiki approach as it is used in the POERUP wiki it runs for the POERUP project, hosted on the Referata wiki farm – http://poerup.referata.com. While there are performance issues for Semantic Wikis as volumes rise, these were not expected to be severe for the relatively low volumes of traffic from the group of OER specialists dedicated enough to make updates. (The situation is quite different for massive *user searches* from across the world – hence our use of the Custom Map Tool for Search in such potential situations.)

There is now a *Create/Edit Open Education Initiative summary* entry page on the POERUP wiki which allows input of the same fields as on the Sero OER Map tool – http://poerup.referata.com/wiki/Form:InitiativeSummary – but behind the scenes builds semantic

information. A bonus of using Semantic Wiki is that geocoding can be done *within the system* and that one can output geocodes to maps. A proof-of-concept of geocoding input is at http://poerup.referata.com/wiki/Form:Location – this has been used for the creation of a number of "city" and "region" pages on the wiki.

These tools are available to all registered users of the POERUP wiki. Interested parties not authorised as POERUP wiki editors but wishing to try this now that the project has ended should contact Paul Bacsich for registration.

To support the user editing phase on the POERUP wiki, the wiki was populated from MongoDB by using the **ImportCSV** Extension available in the wiki, driven by a suitable Template. For an entry point to this data, see http://poerup.referata.com/wiki/Category:Open_Education_Initiatives. At present only a selection of fields from the database is copied across – the full set will be implemented in the exploitation phase of POERUP.

Transfer in the other direction is of course possible via the Query interface or ViewXML – see http://poerup.referata.com/wiki/Special:ViewXML?title=Special%3AViewXML&categories%5BOpen_Education_Initiatives%5D=on.

The data from the current MongoDB database can be seen at the single wiki page http://poerup.referata.com/wiki/Maps_from_POERUP (choice 2) in both Google Maps and Open Street Map format.

5. Plans after the end of POERUP: the exploitation phase

All future plans described below are 'plans', not 'commitments' – in particular the rate and direction of development will largely (but not completely) depend on the range of projects active at the time. However the area of Semantic Wiki and semantic web is an active research topic for Paul Bacsich and he has been developing wiki databases in MediaWiki since 2006.

The general plan of future development of the Map Tools is as follows:

1. Add new fields to the abstract database (first instantiation is always on the Excel version)
2. From then on, for new initiatives, collect data also on these additional fields; for existing initiatives in the database, add the new fields to existing data as updates come in or are required
3. Prototype future maps in Semantic Map Tool or related tools such as Google Map Tools
4. If required for performance or interface reasons (e.g. for mobile platforms), instantiate the upgrade in the Custom Map Tool.

Search functionality will continue to tend to lag behind the addition of new fields.

5.1 First additional implementation after end of POERUP

This is planned to include the following additional search functionality, within the *existing* fields:

1. Search on region (ISO 3166-2) to one level of depth (already implemented on the wiki in terms of region names)
2. Search on language (ISO 639-1) based on the two-letter codes (already implemented on the wiki in terms of language names)
3. Search on ISCED 2011 levels (not yet implemented on the wiki).[2]

[2] This was implemented just as POERUP was completing its reporting phase.

In addition, there will be an editor so that users can recommend *changes to an existing initiative* (in the existing Custom Map Tool a user can *recommend a new initiative*, entering just a few key fields).

The anticipated release date for this is December 2014.

5.2 Second additional implementation after end of POERUP

The specifics will depend substantially on the views of users and the needs of projects. Features that *may* be implemented are:

1. Search on K-levels 1 through 12
2. Search on subject *names*
3. Search on region (ISO 3166-2) to two levels of depth, i.e. subregions also (subregions need to be added)
4. Search on language (ISO 639-3) based on three-letter coding
5. Search on ISCED-F 2013 numeric subject identifiers (ISCED-F field needs to be added)
6. Search on institution names[3]
7. Search on NUTS hierarchy (NUTS-1 and NUTS-2) for the countries where this is relevant – EU and OECD, mainly (NUTS-1 and NUTS-2 fields need to be added)
8. Compound searches – e.g. groups of countries, countries and levels etc (compound searches are already available on the wiki, for the fields currently stored on the wiki).

The anticipated release date for the full set of this functionality is April 2015.

6. Identification of partners for future collaborative work

6.1 In OER and related areas of open education

For future work and projects in this area, Sero, where appropriate in collaboration with its POERUP partner Athabasca University, will reach out into and beyond the current POERUP consortium. Entities that POERUP is already in contact (beyond its own partners and consultants) include UNESCO IIEP, IPTS, OER Asia, OER Africa, the African Virtual University, the Canadian Virtual University, the OER Foundation, and POERUP's own International Advisory Committee (several from the above organisations), augmented by partners in newer projects in particular eMundus, SharedOER, VM-PASS and D-TRANSFORM.

Some sense of the existing coverage of Sero-related individuals in OER can be gained from the *People of POERUP* map at http://bit.ly/peopleofpoerup.

Sero in particular is likely to wish to maintain its focus on "Europe", taken in the wide sense of the European Higher Education Area, and beyond into EMEA and West Asia.

6.2 In online learning more generally

The mapping tools that have been developed are also of considerable value across the whole space of open education, and beyond into the further reaches of Opening up Education: online masters programmes, virtual schools, and virtual mobility.

[3] This was implemented just as POERUP was completing its reporting phase, as a by-product of a separate development path. For the latest tools available see http://poerup.referata.com/wiki/Search and http://poerup.referata.com/wiki/Maps

Sero's other work in e-learning on quality, change management, benchmarking, costs and leadership is already in part supported by wikis (and has been for some years), and in a few cases now by maps. See in particular:

- the ENQA layer of POERUP Map 400 – https://mapsengine.google.com/map/edit?mid=zYG2prGO09jE.kdZ-SJFcZEoM
- 400 Virtual Campuses from Re.ViCa 2007-09 – https://mapsengine.google.com/map/edit?mid=zYG2prGO09jE.kdZ-SJFcZEoM

Two versions of the dynamic Semantic Map on the POERUP Semantic Mediawiki site, using the dynamic version (i.e. incorporating all editor changes) of the Sero OER Map database

With marker clustering

Without marker clustering

Further developments of mapping are expected from Sero in the areas of quality and benchmarking and in new work on virtual institutions.

7. Note: on Linked Open Data in mapping open education

*A small sub-work item was commissioned from Dick Moore on the role of **linked open data**. This took the form not of a static report, but of ongoing advice to and discussion with the work item leader when they were deciding whether to dynamically import or pre-store key datasets required to support the database. The following short section summarises the main decisions taken.*

The OER Maps database and mapping depend on a number of key dataset look-ups:

- ISO 3166-1 alpha-2 for country codes
- ISO 3166-2 for region and subregion codes
- ISO 639-1 2-letter language codes
- ISO 639-2/T alpha-3 3-letter language codes
- mappings of countries to continents and georegions

In each case *reverse mappings* are also needed. For example the reverse mapping for ISO 3166-1 alpha-2 is used to generate canonical country names (a larger range of commonly used country names is allowed for input).

For these it was in the end decided to pre-store all the data. This was done in a small number of additional sheets in the standard data collection Excel spreadsheet. Another advantage of pre-storing is that there are considerable discrepancies between the accepted English name of an entity and the name found in datasets such as ISO 3166 (in particular ISO 3166-2 is notoriously out of synch with common practice, as in some cases is Wikipedia) and we fix these by judicious editing. It might be thought (if only by those with limited experience of data entry) that pop-up lists are a good way to enforce standardisation – they are quite a good way if there are only a few options or where speed of data entry is not an issue. When entering hundreds of entries covering hundreds of regions, entry speed is crucial.

Issues with geocoding

In addition there is the issue of geocoding of institutions, cities, regions and countries. While it is in theory possible to geocode "on the fly" and in some piloting work we have done that, there are license, throughput and performance issues which render that inadvisable as a general approach, and distinct gaps in geocoding coverage. In particular high-profile universities are well geocoded, but many lower-ranking universities in non-English speaking countries are not well covered, and the situation for distance learning providers and new players generally, especially schools and VET providers, is even worse. The example of Virtual Schools in the US using minimal address information was salutary when one checks the details and the manual rework to get it to that level – see https://mapsengine.google.com/map/edit?mid=zYG2prGO09jE.kTnilpzoPwsU – the corresponding example for non-US virtual schools contains too many errors even to load.

Naming issues including Unicode

The need to have a compatible wiki version of the OER database introduces further constraints. In particular it is much easier to load wiki pages if page titles and page data are within the standard ASCII range rather than needing accented or other special characters from Unicode. It is also helpful to wiki use if the use of punctuation in wiki titles is seriously restricted. In that topic area, the usage for university names is notoriously inconsistent, and as long ago as the Re.ViCa project (2007-09) considerable attention was paid to the naming conventions for institutions – see http://virtualcampuses.eu/index.php/Naming. In general terms, the approach of Wikipedia (in our case http://en.wikipedia.org) is far more English-centric than custom and practice in the academic world – thus Wikipedia is not as good a guide in this area as it is in others.

There are further problems caused when one is trying to keep names short and unique on a wiki and yet wishes to avoid the complex disambiguations of full Wikipedia. We resolved these by giving priority in name clashes to countries over regions (so Georgia is in the Caucasus, not in the US) and to cities over regions (so Washington is a city not a state). At this stage we have left the issue of a comprehensive list of abbreviations for institutions as "too hard to handle" in POERUP, though Re.ViCa made a brave attempt – see http://virtualcampuses.eu/index.php/Abbreviations – which could be built on.

Limitations of harvesting

It cannot be stressed enough that *harvesting data* does not *improve data*, and it is wise not to assume that all curation is as good as what we ourselves aspire to. The initial harvest of OpenDOAR into POERUP revealed a large number of issues that required manual intervention: some (like ampersand-coded character codes) may have been infelicities of the XML import to Excel, but most others including institution names in Cyrillic or incorrect geocodes were not. For the final result of our first OpenDOAR harvest see http://poerup.referata.com/wiki/OpenDOAR

Final thoughts

Having taken the initial view that pre-stored data was the best solution, we are in fact in process of revising this approach as we move on beyond institutions to country entries. In related work to POERUP we routinely use a standard set of World Bank data – see for example the report on http://virtualcampuses.eu/images/d/d6/Jordan_HE_online_open_and_distance_learning.pdf – and it is likely that future country-level reporting of the sort piloted at http://poerup.referata.com/wiki/Main_Page#27_August_2014_Leading_countries_for_open_education:_extracts_from_the_POERUP_Open_Education_Initiatives_database_on_the_wiki will use *dynamic* access to World Bank statistics. For numeric data – and provided there is a default value – this is a realistic approach; for textual chunks, less so (as noted above). For other interesting examples of work in this direction see http://emundusatlas.org/socioeconomic.

Harvested OpenDOAR dataset displayed on the POERUP wiki in OpenStreetMap