eIUS Prototype Use Cases[1]

In these examples, underlined elements are based either on the original interview data or on feedback sent by the informants in response to draft versions of the use cases. Key *activity types* are highlighted in *italics* and **relevant ICTs** in **bold**.

Use Case 2 – Applied Econometrics

Narrative

1. Sally, a labour economist, based at Stirling University and with a speciality in the welfare of ethnic minorities is waiting patiently. Three years ago at a conference she met John, an econometrician and they decided in passing to look each other up the next time an opportunity for a new research project arose. Sally is hoping that together they can bring mutually complementary perspectives to each other's research.

2. Arriving early at Southampton University for a face-to-face meeting with John to discuss the writing of a new proposal for their first prospective project together, Sally decides to have a coffee and *check her email* using the wireless network in the university canteen.

3. For the past two weeks, Sally has been *doing some modelling* using an **innovative web portal** developed by John's team. This web interface allows Sally to *perform a kind of statistical modelling previously only accessible to those with an extended knowledge of the Linux command line, high performance computing, and Fortran*. (Sally has never used the command line.)

4. Sally *launches the portal* in order to check over the data she wanted to discuss with John. Because Sally had already signed onto her university's bibliographic and data access system, she does not need to register with the portal or the underlying computing systems it uses because it is also controlled under the UK access management federation for higher education.

---

1 Four use cases were produced for the 'Scoping e-infrastructure' interim report, available online at http://www.eius.ac.uk/scoping/index.xml. They are reproduced here as separate documents to illustrate how researchers in various academic disciplines use e-infrastructure systems in their work

5. *Visualising the study* she had prepared earlier, the results, presented over a map of the UK show the <u>distribution of inequality</u> that appears to validate her own research. However, the model prepared by John does not incorporate a particular <u>macro-economic variable</u> of central importance in her research.

6. John arrives on time and they decide, given the quiet conditions in the canteen, and the fact that Sally already has the portal open on her laptop, to have their meeting there and then, rather than going up to John's office.

7. They *discuss the data* and agree that the existing portal does not incorporate a number of <u>key regional macro-economic variables that would control for regional price differentials</u> (for example, the fact that London is generally more expensive than Scotland). John explains that addressing this would involve incorporating an additional dataset but agrees to *make the changes* necessary within the next couple of weeks.

8. Over the next two months, the portal proves invaluable for Sally for around 80% of the research questions she wants to ask. However, for the remaining 20%, Sally has to ask John to *add bespoke improvements and analysis options* for the model she was originally using. Nevertheless, the portal removes a major bottleneck and allows Sally to make considerable progress in the majority of cases, without any intervention from John.

9. In May they have together collated sufficient data and are ready to *write the results up* in their first academic paper. Sally *selects the subset of the* <u>*archived cases*</u> that is most pertinent for their study of inequality. This is *presented* at a series of UK and European economics conferences and the first article appears in the <u>Scottish Journal of Political Economy.</u>

Relevant ICTs

ICT	Comments
Innovative web portal for labour economics	The GEMEDA[2] demonstrator portal, originally developed under the NCeSS pilot demonstrators project, closely resembles the system described in the use case with the exception that GEMEDA is not controlled under the UK access management federation. GEMEDA requires users to possess three independent security credentials: (1) a GEMEDA service username and pass-phrase, (2) a valid UK e-Science Certificate[3] to access the National Grid Service underpinning the service, and (3) an Athens[4] account to access the datasets exposed by the GEMEDA service. Although the GEMEDA service at the time of writing is still publicly available, lack of funding means it is no longer properly supported as such
High performance computing (HPC) resource underpinning the portal	The name of an actual HPC resource is not given because the researcher interacting with the portal should not need to know which particular HPC resource is used. As stated above, the HPC resource underpinning the GEMEDA portal is the National Grid Service (NGS)
UK Access Management Federation for Education and Research[5]	The federation provides a single solution to access online resources and services for education and research

---

2 pascal.mvc.mcc.ac.uk:9080/gemeda
3 www.grid-support.ac.uk/content/view/23/182/
4 www.athens.ac.uk
5 www.ukfederation.org.uk

Commentary

The research question in this use case was taken directly from the interview data, as was the nature of the collaboration between the applied econometrician and the labour economist. The use case hints that the portal is more of a prototype in nature than a properly supported service. However, the use case is slightly idealistic in the sense that the portal is integrated with the UK Access Management Federation, which at the time of writing, is still in its early stages of becoming established across UK higher education.

Comments by Informant

In this example, the informant who had contributed to the development of the GEMEDA portal, made several changes aimed at making it more believable. He also gave additional contextual information that is included below. The informant suggested GEMEDA not be named within the use case itself while other elements of the use case were for illustration only.

The usual mode of working for the labour economist would be to use a standard proprietary statistical package (for example, Stata[6], SPSS[7], SAS[8]), or a package add-in to perform the modelling required. The labour economist would usually work with secondary data supplied by providers such as the ONS[9] (Office of National Statistics) or the ESRC Data Archive[10] at Essex. She would need to clean the data herself, maintain working copies and manage her modelling and results through package scripts or records of interactive sessions. The labour economist would have a better understanding of the data she works with than the applied econometrician, and the literature and modelling fashions relating to the substantive topic under study. The modelling and the production of the labour economist's results are, however, restricted to that available within her package of choice. The applied econometrician, by contrast, is familiar with the package the labour economist uses and can naively extend it by using simple scripts or macros. However, he can also write his own bespoke modelling code using either specialist matrix programming languages (for example, Gauss[11], Ox[12], Matlab, Octave[13]), statistical languages (for example, R[14]) or third generation programming languages (for example, Fortran, C). He may also be aware of weakness in the algorithms used in the proprietary packages and knows how to parallelise serial code.

---

6 www.stata.com
7 www.spss.com
8 www.sas.com
9 www.statistics.gov.uk
10 www.data-archive.ac.uk
11 www.aptech.com
12 www.doornik.com/ox
13 www.octave.org
14 www.r-project.org

Commenting on step 9, the informant noted that the portal is self-curating and annotating so that details of all the cases Sally has investigated are kept along with the results of the requested analyses. He also made the point that Stirling University does research in applied labour and other microeconomics, one of its professors used to be on the board of the Journal of Political Economy.

Other Editorial Considerations

Element	Usage
Links to direct quotes?	No
Year?	No
Month?	No
Time of day?	No
Location given?	Yes
Real institutions named?	Yes
Real journals named?	Yes
Real conferences named?	No