eIUS Use Cases

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 8 - Crystallography

Narrative

1. Judy is a researcher at a chemistry department at a University in the UK working on new drug molecules. For her research she needs to *determine the structure of those molecules* by means of a method called crystallographic X-ray diffraction, in which a laboratory-based apparatus called diffractometer is used to fire X-ray beams on tiny samples to measure the scattering patterns leading to a model of the molecules.

2. Judy is aware that the **EPSRC UK National Crystallography Service (NCS)** is providing this kind of analysis to support researchers like Judy who are without a direct access to such facilities. Having completed the initial literature review and consolidated her research plan on new drug molecules Judy *uses the service's* ***online form*** to successfully *apply for a one year allocation* on the full structure determination. She then *submits the actual samples via mail* including a printout of the **sample submission form** to provide the NCS with the context of her research and the parameters for the procedure.

3. From time to time Judy *logs in the* ***NCS Interactive Services Portal*** using the secure **authentification certificate** provided by the **NCS.** Here she *checks progress and results of the latest experiments by logging in*, a secure web-based system to access the services. After some time she receives an **automated email** informing her of the first outcome of the structure determination of one of her latest samples. She immediately opens the **Services Portal** in her browser, logges in and *downloads the latest Word file* with results in form of text, tables and figures.

4. That same morning David and George, two of the NCS researchers conducting the experiments and subsequent data collection and full structure analysis are *having a meeting in their shared office to discuss the diffractometer's configuration* for the next procedure involving Judy's latest sample. George then goes to the laboratory to set up the diffractometer for the next run.

5. While the ***raw data*** *is being collected* by the diffractometer and automatically converted into an **ASCII file**, George goes back to his desk. He *opens the* ***SHELX software*** which he often uses to *further process and analyse the* **ASCII files**.

6. Over the next few days George and David work together closely with Judy on the analysis, a process which has the character of collaborative research. They had an *email exchange* the other day with Judy on how to finalise the structure determination analysis to best match her research questions. Having reached a conclusion on how to proceed David *uploads the first final results to the* ***Services Portal*** *in the* ***CIF file format****, which provides a huge amount of additional metadata*.

7. At the same time Judy receives another automated **Services Portal** email notifying her of the newly uploaded **CIF file**. Happy about the result which enables her to continue her research on new drug molecules, she *downloads the file and adds a few metadata annotations*. She then emails the updated file to David and George and suggests a Skype conference call to discuss the best way of *visualising the results*.

8. In preparation for the Skype call George checks and verifies Judy's annotations in the **CIF file** and then *creates some visualisations using the* ***enCIFer tool***. He saves the visualised models of the molecules and sends them via email to David and Judy.

9. In the call a few days later Judy mentions that she has *started a journal article based on the recent research and asks David and George to contribute. She circulates the draft article as a* ***Word document*** *using the track changes feature*. Having published with David and George before she does not need to spend too much time explaining the nature of the collaboration. Usually she is the principal author and provides the research questions and framework for the article, while the NCS researchers contribute paragraphs on the data from the full structure analysis and the procedure itself.

10. A few weeks later the Judy is notified that the article is accepted by the **Crystal Structure Communications Online journal**. Published by the **International Union for Crystallography (IUCr)** the journal besides the text also requires the **CIF result files**. To this end the ***publishing tool publCIF*** *is used to write the final article* directly using the **CIF file** format.

11. Subsequently George and David have stored all data files created during the full structure determination process *in the* **eCrystals repository**, *a free and open database* hosted by the **NCS** to make the data usable to the entire research community.

Relevant ICTs

ICT	Comments
EPSRC UK National Crystallography Service (NCS)[1]	The NCS primarily uses laboratory-based facilities located in the Chemical Crystallography Laboratory at the School of Chemistry at University of Southampton. The service offers 1) data collection only (with the user doing the further analysis) and 2) full structure determination to researchers from the whole of the UK. To this end a pair of diffractometers is used, a measuring instrument emitting X-ray beams to analyse the scattering pattern on crystals to find geometry or shape of molecules.
NCS Interactive Services Portal[2]	The NCS provides an electronic workflow for its users including an online application form, downloadable documents, a web interface, certificates for user authentification and an underlying SQL database. The NCS Interactive Services Portal was originally developed within the EPSRC-funded testbed project CombeChem[3].
Raw binary data	This form of proprietary data is created directly through the crystal structure determination process and has to be processed into a standard ASCII file for further analysis. The whole of the raw data produced by the NCS is archived in the large object Atlas data store at the STFC Rutherford Appleton Laboratory (RAL).
ASCII files	An ASCII file is a plain text file which can be read by every standard editor or word processor.
SHELX[4]	SHELX is the most frequently used suite of programs for crystal structure determination.
Crystallographic Information File (CIF)	A CIF file is the (final) result file representing information from crystallographic analysis in a standard text format, enabling the use of diverse metadata. Its scientific use in the domain was agreed on in 1992/93 under the agenda of the International Union for Crystallography (IUCr).

1 http://www.ncs.chem.soton.ac.uk/
2 http://interact.xservice.soton.ac.uk/index.php
3 http://www.combechem.org/
4 http://www.ncs.chem.soton.ac.uk/data_coll.htm & http://shelx.uni-ac.gwdg.de/SHELX/

enCIFer[5]	enCIFer is a commonly used visualisation programme used in Crystallography (another widely used tool is Mercury[6]) to visualise CIF result files, offered by the Cambridge Crystallographic Data Centre (CCDC)[7], a self-financing non-profit organisation with the remit to generally promote the science of chemistry and crystallography. It also hosts the Cambridge Structural Database (CSD), an international repository of small-molecule crystal structures and offers other services and tools.
IUCr journals[8], e.g. the Crystal Structure Communications Online journal	A series of online journals for crystallographers is published by the International Union for Crystallography (IUCr). Those kind of journals nowadays require the results file as well. Papers can actually be written in the CIF file format, i.e. IUCr provides tools to render the final paper based on those files, e.g. publCIF (see below).
publCIF[9]	Free software offered by the IUCr to edit and preview a publication based on the CIF format. Changes can be made in the word file or the CIF file and are converted in either direction.
eCrystals repository[10]	Under the label of Open Science eCrystals is an open and free archive for crystal structures data (excluding raw images), their metadata and references hosted by the EPSRC UK National Crystallography Service and the Southampton Chemical Crystallography Group. Its aim is to provide as much data as possible to the community including datasets of experiments on the way to the final result. E.g. it can take 50 steps to come to this final result, but only No. 50 is conventionally published, meaning the other 49 are lost with nobody being able to learn and benefit from those.

5 http://www.ccdc.cam.ac.uk/free_services/encifer/
6 http://www.ccdc.cam.ac.uk/free_services/mercury/
7 http://www.ccdc.cam.ac.uk/
8 http://www.iucr.org/
9 http://journals.iucr.org/services/cif/publcif/
10 http://ecrystals.chem.soton.ac.uk/

Commentary

The story is based on the interviews with two NCS researchers – here called David and George – which provided enough significant information to invent the third protagonist Judy as a user of the service.

The interviewees emphasised the importance of the experience and individual expertise of each researcher conducting the experiments for data collection and full structure determination and analysing the data: “Crystallography more than anything is all about experience” and the difference in “experience is quite noticeable when you get a really difficult problem” and “it requires well three years of training at least plus any postdoc experience”. At the same time the expertise and background of each NCS researcher is of meaning as well the assigned jobs “on our previous experience of what areas we’ve previously worked in”.

In the context of the eCrystals repository but also looking at the sharing of data in the Crystallography community the interviewees stressed their Open Science perspective: “Ultimately my goal is all about streamlining the process of getting data into these databases, whether it’s through streamlining the journal publication process or whether it’s just fast tracking straight to the databases; the fact is, the experiment’s been done, that piece of data should be in that database and at least eighty percent of it isn't”, for two main reasons. 1) “I generate more of these things than I can possibly write up in journal articles” and 2) collaborators sometimes do not want the results to go out, e.g. because they did not achieve the desired result (and might still be in the process of getting it right).

Other Editorial Considerations

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