Post JRC Workshop – RD-ACTION WP6 analysis of issues concerning the ERN Platform for clinical patient management and the EU Platform for RD Registration

We encountered a certain level of platform confusion in the JRC workshop, which is not perhaps surprising. People are uncertain as to what each platform will be able to do, and what it will not. And furthermore, it is difficult at present to conceive of the *degree* of interaction between the two, from the outside perspective at least, if we accept that no data will be conveyed to the JRC from the ERNs (the only sustainable registries hosted in Ispra will be the SCPE and EUROCAT registries). The two platforms are summarised below (based on Jarek's synthesis at the workshop - he provided clear explanations of the different platforms) to support focused discussions.

1. One platform is the ERN IT platform for exchanging clinical data (this is the SaaS). When a patient needs to 'enter' the ERN for some sort of multi-person virtual review, they will have to provide consent for their data to be shared in this way for care. They will also be asked if their data can be reused for research (precise wording to be agreed by the EC). If they consent to latter, the Tender specifications state that the platform will retain, pseudonymize and store this data, as reported in the RD-ACTION document '*What do Coordinators Require from an IT Platform?*'

	The cases reviewed by the ERN in collaborative virtual consultations should afterwards be retained in a case record repository within or accessible via the IT Platform	**(xiv) encrypts and stores the data;** **(xv) pseudonymises patient data for sharing, use in clinical decision making tools, protocols, guidelines, case library or research;** **(xvi) hosts the data storage within EU borders and ensures that the hosting is single-tenant with stable, fast and easy data storage and retrieval, back-up and recovery;** **(xvii) ensures that hosting is redundant at both the database and application server level;**

2. The other platform is the EU Platform for RD Registration (sometimes called the JRC platform, although this is not the official name). This has two main pillars:
    a. one concerns data collection for EUROCAT and SCPE; however, this is not directly relevant for the ERNs as we have been told JRC will not host data for other registries (although the lessons learned from these ambitious ventures surely ARE relevant).
    b. the second pillar is the interoperability platform of JRC. This platform seems to be mainly concerned with setting standards for RD registration, providing tools (whether recommending software for establishing a registry, promoting methodological guidelines, helping to increase the visibility of registries –and utility of the data- through ID cards or a registry or registries, etc.)[1]

[1] In terms of what the two platforms can offer ERNs concerning registries, there are probably two broad and different scenarios here:

1. If an ERN desires a *new* registry, the JRC could provide software and guidance/tools for this OR it *could* possibly be provided as part of the ERN IT platform, if the vendor has particular expertise in registries.
2. If the ERNs have many registries already and do not wish to 'begin again', they will likely need interoperability tools, ways to make data searchable/queryable. Again, the JRC could provide something here (e.g. through the API route, perhaps).

Perhaps the fact that the ERN IT Platform (SaaS) will need to be able to store data for secondary purposes has inadvertently created some confusion. As above, a registry –at least in the *traditional* sense of the word- is different to a set of data elements collected and shared in a virtual consultation. A registry asks a specific question, and is typically purpose-specific (a few thoughts below…) However, the data collected from patients and shared via this SaaS platform for care WILL be valuable and should not simply be destroyed (unless the patient so wishes, of course). The question then becomes, **what do we do with that data, once encrypted, stored and pseudonymized as defined in the Tender specs?** A couple of possible options seem to present themselves at this stage:

a) Could this data somehow '*feed*' a registry? To answer this, it is necessary to examine the 'overlap' between data captured in a registry and the sort of information people wish to share in virtual patient consultations. The level of complexity here would seem to depend on whether one intends feeding data to:
    i. An 'unconnected' registry, i.e. a registry already operating, which is hosted by an entirely different IT system. This sounds very challenging, certainly to do automatically, as data fields would differ, and there would surely be legal complexities
    ii. A new registry created by the vendors providing the SaaS ERN platform, which was built specifically for this sort of purpose.

OR

b) Could this data remain stored in this platform to be searched 'or queried' in its own right, as it is?

Option b) at least seems very feasible and logical at this stage. Using appropriate data interoperability standards, one could presumably render this information queryable for future use by the ERN community; however, it may be necessary to think more about the real value of the sort of data that will be captured and held. For instance, the SaaS platform *might* provide valuable information on genotype/phenotype correlation, on the value of different therapeutic approaches, for trial recruitment or other purposes. At the very least, this information held as a sort of 'record' by the SaaS would be important for the monitoring and evaluation of the Networks, surely – this will help the coordinators to assess the number of patients reviewed by the ERN per se, and could yield information on outcomes etc.

But at this point, it is essential to clarify what might be the utility of this patient data, once a virtual 'consultation' has occurred, bearing in mind the promised capabilities of the SaaS platform as defined above, and considering always the scarcity and thus value of rare disease data.

The JRC workshop discussions clarified the fact that CDEs such as those defined by EPIRARE, RD-Connect *et. al.* were created for registries. When patient cases are 'referred' to the ERN for consultation, coordinators do not envisage sharing a registry dataset between consulting experts. Having accepted that a clinical dataset cannot 'become' a registry in and of itself, a crucial question remains: **'what data do people plan/wish/need to exchange for virtual care?'** If the clinical

information required for each ERN is so different as to render harmonisation impossible, the logical approach is to clearly agree the data fields, the types of information. For instance:

- Demographic data – all ERNs will need to collect some demographic data on the patient, and this should be captured in the same way (e.g. format for DOB must be standardised). At the September RD-ACTION workshop on *sharing data for virtual care in the ERN framework*, the concept of the GUID (Global Unique Identifier)/ PUID (Patient Unique Identifier) was raised.[2] It was proposed at the time that for ERNs to be able to participate in these international RD research efforts in future, they should each capture agreed elements of patient identifiable information (e.g. First, middle and surname as on birth certificate; Date of birth; City of birth as on birth certificate)
- Diagnosis - if the patient has received a diagnosis, this should be coded using the ORPHAnumber, as the most robust and granular coding system for RD. To enable this information to be machine readable (i.e. semantically interoperable), an ontology - the Orphanet RD Ontology, ORDO) - should be used. It may be useful to indicate the degree of certainty of this diagnosis (e.g. is it confirmed or simply suspected?)
- Clinical presentation – to make this information useful for streamlined virtual consultations, but also for re-use, **the manner in which clinical information is captured requires some forethought here**.
    - Clinicians typically prefer to capture information on their patients in free text, making observations on their observed phenotype. If the SaaS opts for this approach, the data will only be useful and searchable electronically if the terms used are agreed HPO terms. Systems such as the Patient Archive allow users to write their notes as they would usually, and it automatically translates terms to an agreed HPO term (e.g. a clinical could write microcephaly or nanocephaly or small head, and it would autosuggest an agreed single term to appear in the final description).
    - An alternative to capturing phenotype of the patient is to record their symptoms and presentation using a form-type interface, in which clinicians are guided as to what items of information to provide. From the presentations at our September workshop, we learned that determining the particular information one wishes to see when reviewing a patient virtually is very important in organising efficient consultations but also in maximising the value of the data later, for analysis. If a structured form is preferred, again it would be important to have predictive HPO-based tools to support interoperability of data for reuse. A PhenoTips-type functionality would be preferable.
    - Perhaps a combination would be best, at least to begin with – the items that ERNs know they will always wish to collect and record when reviewing a patient can be put into a form structure, and there could also be a free text box for clinical notes (again, the functionality of tools such as the above would be hugely beneficial for RD)
- Non-textual data will also obviously need to be uploaded/attached/linked to the patient record. From the survey of Coordinator IT needs, people reported the need for MRI, PET and CT scans, ECG, x-rays etc. (so a DICOM viewer will be necessary). Others added that ideally,

[2] This is a concept being pursued by the global RD research community - the idea is that by all studies/registries etc. always collecting the same few items of patient identifiable information, the same patient will always receive the same pseudonym (e.g. a string of letters or numbers/code) when that PII is entered into a system/server to generate identifiers. In this way, disparate data relating to the same patient (e.g. data held in a registry or in different registries, biosample data etc.) will be linkable.

the system will be able to 'import' files directly from the systems used routinely by the HCPs, such as PACS, LIS, RIS etc.

Rare Disease Registries and ERNs – a few thoughts

WHO defines a registry as "a file of documents containing uniform information about individual persons, collected in a systematic and comprehensive way, in order to serve a pre-determined scientific, clinical or policy purpose".

Registries can serve many important purposes in the rare disease field; for example:

- By collecting data over a long period of time, registries can elucidate the **natural history** of a disease (i.e. how the symptoms develop and progress, what the prognosis might be, etc.);
- Registries can increase understanding of the **epidemiology** of the disease: epidemiology encompasses the *cause* of the condition (**aetiology**) but also its impact on any given population (if the registry covers whole populations, researchers can more accurately calculate the incidence and prevalence of disease). Such epidemiological information is very valuable in assessing disease threats and informing the appropriate planning of health services;
- Registries can demonstrate the **efficacy of different clinical management, diagnostic and therapeutic options,** presuming information on treatment regime and clinical outcomes is captured. For instance, the use of corticosteroids in neuromuscular patients can be tracked against the degree of ambulation and mobility. The relative impact of different regimes of enzyme replacement therapy for patients with inherited metabolic diseases can be assessed with reference to liver and spleen volume, for instance;
- Registries -if established appropriately- can support the **post-marketing surveillance** of approved/conditionally approved orphan medicinal products. Increasingly, the safety and efficacy of medicinal products for rare diseases are granted less-traditional (i.e. 'adaptive') pathways to marketing authorisation, in which a drug may be conditionally approved for use based upon a relatively low volume of trial data (often unavoidable in the rare disease field), on the understanding that high-quality robust data will be captured for each patient prescribed that drug in the years to come.
- The correlation between certain genetic mutations and corresponding clinical presentation (phenotype) may be elucidated by registry data. Sometimes patients with the same condition and the same genetic mutation exhibit very different symptoms and experience the disease with varying severity: only by capturing this information routinely and robustly are researchers better able to understand rare conditions and their prognoses by correlating patients' genotypes and phenotypes (in other words, understanding how different combinations of genetic anomalies result in particular clinical presentations).
- Registries are significant enablers of clinical research, for instance by supporting an assessment of the feasibility of conducting a trail in the first place, and later by facilitating the recruitment of patients. This is particularly useful when registries record an accurate genetic diagnosis (i.e. they stipulate the particular mutation responsible for causing the condition). As medicines and interventions become more personalised, clinical trials often target a specific mutation and therefore need to recruit a particular sub-set of patients. The

existence of detailed genotypic information enables a sponsor to assess the number of trial participants they could potentially recruit, and where they are based. This sort of information is critical in supporting the pharmaceutical industry and academic communities to drive forward much-needed clinical research in the rare disease field.

(More information on types of RD registries can be found in this comprehensive report of the RDTF)

To attempt to increase interoperability of the existing and future registries for RD, experts have sought to promote use of RD-appropriate (i.e. the most RD 'sensitive') ontologies (such as ORDO and HPO) and also common data elements/core datasets. The rationale for the latter is that registries intended for the same purpose (e.g. trial-readiness) will all capture the same relevant data items, which will support aggregate data pooling and analysis, to enhance current knowledge and expertise and drive forwards progress. Examples of such outputs include the CDEs or minimum datasets generated by EPIRARE[3], RD-CONNECT[4] and the EUCERD Joint Action[5] (see also the US GRDR CDEs[6]). **It is important, for the future of the EU Platform for RD registration, to determine once and for all whether there *is* in fact value in registries agreeing common data elements in the field of rare diseases, and if yes, at which level.**

It seems undisputed that there is a major benefit to agreeing data fields for specific rare diseases/groups of diseases – many examples exist to demonstrate the added value of this approach. For instance:

- In the case of Duchenne Muscular Dystrophy, national registries establish their own datasets, collecting whatever information they wish, but they all agree to capture the same mandatory TREAT-NMD data items[7], which are uploaded to the global TREAT-NMD database. These mandatory items include demographic data, genetic diagnosis, ambulation status, wheelchair use, etc. Then there are a number of agreed 'highly recommended' items which most of the individual DMD registries also capture. This approach allows the data to be pooled for purposes such as clinical trial recruitment and assessing the impact of steroid use.
- The EURO-WABB project established a European registry for the so-called WABB conditions (Wolfram, Alström and Bardet-Biedl Syndromes) by agreeing a common dataset of 44 core items for all rare diabetic syndromes. This was accompanied by an extended dataset, comprising in total 370 possible fields for detailed phenotype information.
- In the E-IMD initiative (European registry and network for Intoxication type Metabolic Diseases), the partners agreed a set of core data elements which would be applicable to all Organic Acidurias and Urea Cycle Defects, to provide a robust baseline for comparable data on all patients in the registry: more specific data items were then defined for each particular disease.
- When creating the International Niemann-Pick Registry, the partners agreed one core dataset for NP Type C and another for NPA/NPB.

[3] http://www.epirare.eu/_down/del/D9.3_ProposalforCDE_FINAL.pdf
[4] http://rd-connect.eu/rdcon/files/RD-Connect_CDEs_May_2016.pdf
[5] http://www.eucerd.eu/wp-content/uploads/2015/03/WP8_Registries_MDS.pdf
[6] https://ncats.nih.gov/grdr/cdes
[7] http://www.treat-nmd.eu/downloads/file/registries_toolkit/DMD_core_dataset_May2013.pdf

There is thus surely value in European Reference Networks agreeing common data elements for the registries in their field. However, is there added value in attempting to define a common data set for *all* registries in the RD field, including those established by/ used by ERNs? Past discussions on the topic of a common data set have concluded that the data items common to all RD would be very limited. Some of the proposed CDEs/Minimum Datasets referenced above seek to define mandatory data items depending on the purpose of the registry. These discussions surfaced at the JRC workshop in November and perhaps need to be revisited at this time, if there is a prospect of one or another common/minimum datasets forming part of the expected 'tool-kit' for registries to be provided by the JRC.

Another important issue to clarify is precisely what constitute the ERN 'framework'? In past discussions we attempted to determine if only information from patients 'referred' to the virtual care of the ERN will have their data entered to the SaaS platform. Is this indeed the case? If so, it is the RD-ACTION perspective (certainly from Newcastle team, and also from Yann le Cam's presentation on 30th November) that ERNs should attempt to expand enrolment of patients in registries (wherever they are hosted). As we wrote in our paper *What do Coordinators Require from an ERN IT Platform*? "The more (high quality and comparable) data one can aggregate, the greater the potential for knowledge generation. Therefore, when contemplating the storage and re-use of anonymised patient data, are we thinking only of data from those patients referred for shared care within the ERN (i.e. via a virtual/MDT review) or of all the patients with a particular condition who visit the HCPs within a given ERN?" Some ERNs have recommended to their HCP members that increasing registration of patients visiting their centres on a day-to-day basis should be a priority. If ERNs are indeed intended to be game changers, it seems logical that any patient seen by any HCP in an ERN should be offered the opportunity to enrol in a relevant registry. But again, this perhaps needs to be discussed with the ANCs.