Clinical Data Scientist – what it is (not)

Michaela Jahn, F. Hoffmann-La Roche, Clinical Data Management Early Development, Basel, Switzerland

ABSTRACT
Within clinical data management, more and more companies are talking about becoming a Clinical data scientist Department. But, what in fact is a Clinical Data Scientist? Is it a renaming exercise – all Data Managers become Clinical Data Scientists? What is the difference between a Clinical Data Manager and a Clinical Data Scientist? Without providing the ultimate correct answer, this paper contrasts the roles of a Clinical Data Manager and the role of a Clinical Data Scientist. As well we discuss how the transition from Clinical Data Manager to Clinical data scientist can be supported; outlining the challenges and sharing success stories.

INTRODUCTION
Change is the one constant element in the drug development working environment. For quite some time we were looking at the evolution of clinical data management. On that path we covered outsourcing, offshoring, big data, data sharing, patient-centric data. What does the latest edition ‘Clinical Data Scientist’ add? In the following, we describe a personal view on what the element of a clinical data scientist can add to the discussion of the evolution in clinical data management. Without providing the ultimate correct answer, this paper contrasts the role of a Data Manager and the role of a Clinical data scientist.

EVOLUTION IN CLINICAL DATA MANAGEMENT
Looking back, data management was a very task driven organization (Figure 1). Those times changed and data management reacted to new data types (for example patient reported outcomes), new technologies (more electronic transmitted data). We were able to manage almost whatever was provided to us. Still, we were in a reactive mode. Over time, teams became more demanding. Data were no longer regarded as owed by biometrics functions like statistics or data management. Data availability and data access was demanded by the study teams. Data Management reacted with data access and training like on data models, standards and the shared responsibility.

Figure 1 Data Management - Past

Now we are facing the topic of clinical data science. What does this entail?

DEFINITION CLINICAL DATA SCIENTIST
Before talking about clinical data science we should investigate the definition of a ‘Clinical Data Scientist’. Is there an agreed definition? Some random examples:
Data Science is the extraction of knowledge from large volumes of data that are structured or unstructured, which is a continuation of the field data mining and predictive analytics, also known as knowledge discovery and data mining (KDD) – Wikipedia (search term 'Clinical Data Scientist').

Data Science: applying the team's diverse informatics and analytical capabilities to retrieve and analyse data to support drug project decision making, drug and platform development. Data Science capabilities include Bioinformatics, Imaging Informatics, Biostatistics, Data integration and visualization, Text-mining, Information Science – Roche pre-clinical and early clinical development.

Clinical data scientist has a comprehensive knowledge of all areas pertaining to the management of data, data delivery, understanding protocols, is able to interpret clinical study data, and the technologies needed/used on clinical studies from start-up to completion – Roche early clinical development data management.

This is just a small excerpt of available definitions. But it already illustrates the broad use of the term ‘clinical data scientist’. Common features in the above definitions are related to the management of information or knowledge. The data are linked to interpretation and meaning. The people in charge of the data/information/knowledge need to understand more than just managing data. It’s about managing related information and knowledge.

In summary, there is no single definition for the clinical data scientist. For the remainder of the paper, we will focus on the last of the above definitions and illustrated the transformation from a data management department to a data science department.

CONTRASTING CLINICAL DATA MANAGER AND CLINICAL DATA SCIENTIST

SKILL SET OF A CLINICAL DATA MANAGER

The skills and competencies relevant in the traditional data management

- Timeline management
- CRF/eCRF design
- Review query log
- Raise manual queries
- Review answers to manual queries
- Handle science queries
- Clarify errors
- Check of process regarding coding of terms
- Arrange for loading of electronically loaded data
- Reconcile serious adverse event between clinical data base and drug safety database
- Reconcile non-eCRF and CRF data
- Generate dose escalation snapshots
- Amend the database when necessary
- Train sites/monitors for study conduct
- Provide metrics to study management teams
- Provide data listings to vendors
- Prepare for database closure

Although regarded as traditional tasks, the above tasks need to be done in orders to get to the ultimate goal of providing clean data to be extracted and handed over to other functions such as biostatistics, pharmacology, clinical science, and modeling and simulation. There is a need to perform the above tasks.

PERCEPTION OF A CLINICAL DATA MANAGER

In numerous pharmaceutical companies, the data management department works on and is limited to the tasks outlined above. Within study management teams, data management and the study data managers are perceived as being at the receiving end of information - being asked to perform the work and acting as a service provider to the organization. As such, the role of data management is frequently named first when it comes to the identification of transactional tasks that could be outsourced or off-shored.

This trend indicates the common perception that the data manager works in isolation, provides a routine service, and is not an integral part of clinical teams.

We dispute this perception and argue that the skill set available in data management can provide much more, and the demands and the opportunities are real.

ATTRIBUTES OF A CLINICAL DATA SCIENTIST

Clinical data scientists need to be regarded as equal partners in the study management team. In order to achieve this,
the data management responsible person needs to resemble key attributes:

- Have a clear understanding of protocols, their structure, primary and secondary endpoints and what it means in terms of being able to accurately collect and extract the data. Reach out to partners like clinical scientist, biostatistician in case of questions
- Oversee study milestones and what is needed for data delivery
- Conduct a risk assessment in regards of data, e.g. what needs to be clean for a therapeutic area and what data can be left as is
- Understand the basic needs of statistics and programming
- Support standards
- Understand the basics of the disease area
- Oversee external service providers
- Adapt to new technologies
- Help clinical scientist to understand the data modeling and to explore the data

MOVING FROM TRADITIONAL DATA MANAGER TO CLINICAL DATA SCIENTIST

In the following, we describe the path we took. This path does not present the only possible route. Multiple paths can be followed on the journey to become a clinical data scientist.

By addressing the attributes outlined above in different forms of communications, trainings and workshops we set expectations to the members of the department.

CLEAR UNDERSTANDING OF PROTOCOLS

In order to work with data/information and knowledge, the clinical data scientist needs to understand the study protocol. Curiosity and willingness to learn by asking questions is essential for the clinical data scientist.

We fostered this behavior by partnering with other functions who presented introductory learning sessions re their area of expertise. This included:

- Practical introduction to statistics for non-statisticians - basis understanding of the statistics needed for studies and how is translates from protocol text to analysis
- Clinical Pharmacology – Basic understanding of the Clinical Pharmacologists role in study management teams
- Pharmacometrics – How to set up and perform the analysis and which data are considered
- Drug Safety – Overview of the role and the tasks
- Programming – Introduction to SDTM

With the introductions of different partnering functions and linking the protocol text to actual tasks performed by the partnering functions, an open dialog started. With the invitation and encouragement to ask questions during the presentation but as well on study management teams, initial hurdles started to be taken.

In dedicated workshops small groups of data managers were guided through protocol review. Different protocols were designed to support the protocol review. Mock-up protocols with errors and inconsistencies were designed to practice the protocol review.

OVERSEE STUDY MILESTONES

With adaptive designs, interim readouts, safety data monitoring, a clinical study presents multiple milestones. Each of these reporting events needs to be carefully planned and all required activities need to be performed. With the increased complexity of clinical studies, the planning has become more complex. Very often simple tools are no longer good enough to support the data management department. To act as clinical data scientist, the milestone planning is essential for the function itself, but as well to communicate within the study management teams.

The knowledge of project management techniques and tools allows to present data management as a competent partner to the study management teams.

We offered to the data managers project management training. Depending on previous knowledge, various learnings were made and applied in the day to day activities.

CONDUCT RISK ASSESSMENT

A clear understanding of the study deliverables is a pre-condition for a solid risk assessment. The clinical data scientist must understand which data points are included in the most important analyses. For those data points, additional quality measures need to be set up to ensure completeness and accuracy. We implemented a process to specify the key data points most relevant for the analyses at different reporting events (data delivery expectations). Also the expectations regarding level of data cleanness were discussed and specified.
OVERSEE EXTERNAL SERVICE PROVIDERS
The qualities to work with external service providers are similar but not identical to the in-house study model. Building relations, manage expectations and identify the shared purpose are essential for a clinical data scientist. Besides guidance documentations, we highly recommend to set up early on dedicated data management meetings (best face-to-face). At the meeting, data management staff from the sponsor and from the service provider working on the study should get together to plan the study, its deliverables, milestones. But at the same time, expectation management and clear splitting of responsibilities is discussed.
Further information can be found in the paper ‘Ensuring Data Quality on Outsourced Studies - The Role of the Data Scientist, presented by Sarah Malbon and Vikki Horton at the Phuse Conference 2015.

ADAPT TO NEW TECHNOLOGIES
New data types are entering clinical studies regularly. The clinical data scientist needs to stay on top of the development and assess the impact of the new technologies on clinical data management. Two examples to illustrate this:
Clinical images have become part of the data and information that is managed by a clinical data scientist. Basic information regarding imaging techniques need to be acquired to manage the related data, information and knowledge.
Especially in the central nervous system area, mobile devices are considered to for example assess patient’s mobility. What kind of data will that bring to a data management department? A clinical data scientist needs to evaluate the technology, liaise with experts and be willing to explore new data types.

SHARE UNDERSTANDING OF DATA MODELLING AND DATA EXPLORATION
More and more clinical scientists would like to have rapid access to data collected in studies. With technology and new tools, this is possible. But in order to explore and interpret data, the clinical scientist needs to understand the data structure and the data modelling that is done, before the data can be explored.
Here, the clinical data scientist has a teaching role. We support clinical data scientists to start the dialog with their colleagues. The joint knowledge of data linked with information and knowledge results in a powerful combination.

CONCLUSION
The evolution of the clinical data management function to become more like a clinical data scientist function is a multifaceted transformation. It requires technical and soft skills broader than in the past. Therefore the expectations on the job are changing.
The adoption of the clinical data scientist role is not just a renaming exercise. But the transformation is essential for the future of clinical data management within the pharmaceutical industry.

ACKNOWLEDGMENTS
I would like to acknowledge all my colleagues in pRED Clinical Data Management who bought into the journey of becoming data scientist. Special thanks go to Michelle Poskonka and her team for paving the way and helping with the content of this presentation.

CONTACT INFORMATION
Your comments and questions are valued and encouraged. Contact the author at:
   Michaela Jahn
   Global Head pRED Clinical Data Management
   F. Hoffmann-La Roche
   Bldg 663/2111
   4070 Basel
   Switzerland
   michaela.jahn@roche.com

Brand and product names are trademarks of their respective companies.