Royal Botanic Gardens Kew

Section 1: Job Description

Job Title: Data Scientist: *Plants for Health*	**Post No:**
Band: D	**Job family: Science & Scientific Information**
Directorate: Science	**Section/Dept: Digital Revolution/Research**
No. of Direct Reports: 0	**No. of Indirect Reports:** 0
Reports to (Position): Digital Resources Manager: *Plants for Health*	

Job Purpose:

To deliver RBG Kew's mission by leading successful implementation of *"Plants for Health"*: a key deliverable of Kew's Science Strategy (2021-2030) contributing primarily to Priority 3: "Digital Revolution".

The Data Scientist will use their experience and expertise in data mining and manipulation (including AI & ML) to locate sources of health data relating to plants and plant products and devise the means to extract, effectively and efficiently, those data relevant for inclusion within the *Plants for Health Resource*. The post-holder will need to become familiar with the research and regulatory literature in diverse health domains and to develop tools for data mining for their own use and that of other project staff and collaborators.

The post-holder will assist the Digital Resources Manager in implementing quality control, data maintenance and analytical reports. They will further liaise, as required by the Digital Resources Manager, with the data managers from partner agencies which either hold relevant digital resources or employ software tools (e.g. AI) of potential use to *Plants for Health*.

Job Context: 1

Plants for Health is a new Kew initiative running initially for 4 years supported by funding from the Wellcome Trust.

Plants for Health builds and relies upon the successful "*Medicinal Plant Names Services (MPNS)*". MPNS has become the default global reference for herbal medicines and plant-derived drugs being employed in drug regulation, industry, and natural products research. *Plants for Health* will expand the products and data classes to address the information needs of a significantly wider set of audiences including those working with nutritional supplements, beverages, allergens, toxins, and cosmetics.

Seven global health institutions are partnering Kew to build *Plants for Health* and are contributing data, domain expertise, software and methodologies as well as permitting access to their own user audiences and collaborative networks. A wider set of industry experts will advise and guide the project.

This post will report to the project's Digital Resources Manager. The project team at Kew will consist of five full-time staff who will compile and curate data, build and deploy the digital products to meet defined audience needs. Existing MPNS work streams and projects will continue to run in parallel to *Plants for Health* in the short term although we anticipate that MPNS will eventually be subsumed into the broader initiative.

Accountabilities:

1.	**Explore the use of AI and other digital tools to increase the efficacy of data capture**. Adopt, build and document a suite of alternative methods and tools for locating, assessing, mining and capturing data from reference works, text corpora and structured databases. Liaise with Kew's partners and advisors with expertise and tools of potential use within the project.	*15%*
2.	**Capture datasets from health sources using agreed tools and techniques** as effectively and efficiently as possible to contribute the bulk of the health-related data within *Plants for Health*. Undertake digitisation of key reference documents internally or procure and manage 3rd party organisations to digitise priority references.	*35%*
3.	**Work with the data managers and IT staff of external partners** under supervision of the Digital Resources Manager to: a) facilitate the two-way exchange of data with *Plants for Health* by designing and implementing effective mechanisms. b) document partners' data needs to be able to contribute to the definition and prioritisation of requirements for digital services to be developed by *Plants for Health*. Support wider communication with project stakeholders and users.	*15%*
4.	**Support the Digital Resources Manager in designing and implementing data quality controls and tools**. Contribute to the *Plants for Health* catalogue of potential data sources (describing plant-based health products from among diverse health disciplines) by locating and adding potential sources that might be mined and captured using the suite of developed tools. Carry out data analyses to monitor data quality, database growth and identify gaps in our data, generating reports and statistics as required by the Digital Resources Manager.	5%
5.	**Manipulation & import of health datasets:** undertake data transformation and cleaning exercises to organise and standardise captured data ready for incorporation into the *Plants for Health* Resource. Map plant and fungal scientific names lists to Kew's core nomenclatural and taxonomic references employing agreed protocols and tools. Follow established protocols to check and import cleaned and matched data into the *Plants for Health* Resource.	25%
6.	**Co-author research papers** relating to the digital resource, information products, methodologies and wider relevance and impact to human health. **Contribute to the project and wider "Digital Revolution" teams**; attend meetings, report progress, give presentations, and support colleagues working on other biodiversity information resources as required.	5%

Management of Resources

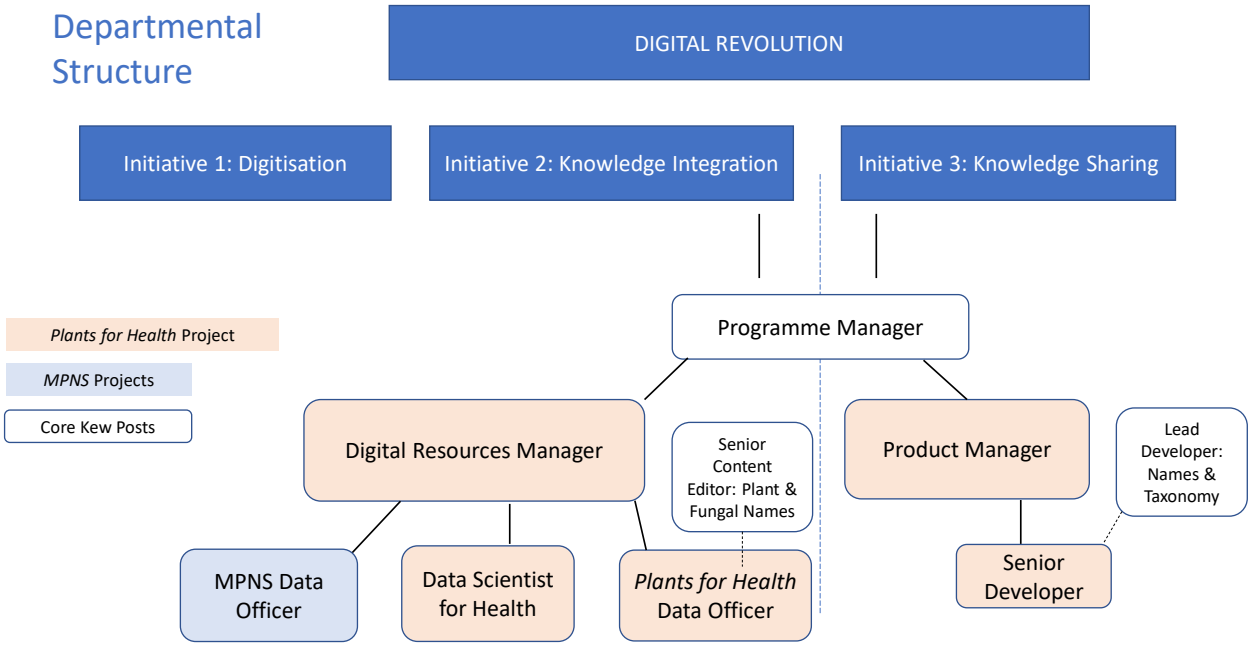
Financial Responsibilities*:* All staff are responsible for management of their own time and resources.

1.	Effective use of project resources for procuring 3rd party digitisation of selected references.
2.	

Management of People

1.	Mentoring and training of junior staff and volunteers in the use of data capture and data analysis tools and techniques
2.	

Position within the Team



Section 2: Person Specification

Education and Experience

Essential
A degree in a relevant subject or evidence of having equivalent experience of managing complex or scientific data within a scientific or professional organisation
Formal qualification or relevant experience in data management and transformation including practical use and proficiency in several of the following: SQL, XML, R, Python and mark-up technologies
Demonstrable experience of the capture, curation and deployment of complex scientific data
Demonstrable experience with 'Relational' database design and data modelling
Significant experience working with information management and data design
Demonstrable experience with a diversity of techniques and tools for mining data and of having used them to detect and/or extract data from complex datasets or text resources
Demonstrable experience with health or biodiversity informatics (or equivalent scientific field) and relevant global data standards
Familiarity with research paradigm
Desirable
Experience with health data management and of health digital services in the UK or globally including pharmaceutical industry and regulation, drug, food or other product safety.
Experience of having liaised effectively with external partners, facilitating workshops and user training, trialling data management tools and partner data exchanges.
Evidence of creative problem solving.
Evidence of producing technical and data reports intelligible to managers, users and 3rd parties.
Experience of RBG Kew's existing information products, databases and data management practices
Experience of sharing, refreshing and re-use of data through API interfaces
Evidence of publishing scientific papers relevant to data analysis or management.
Formal training or experience of workshop facilitation techniques
An understanding of the complexity of scientific plant names and familiarity with how non-botanists misuse them. Evidence of designing effective techniques for matching scientific plant names against digital reference resources such as those held by Kew.
Familiarity with and interest in the Natural Health Products industry and research

Job Specific Competencies

Essential
An understanding of data quality assurance and the significance of appropriate data structures
Highly motivated, energetic, and inquisitive.
Well organised, able to work alone, plan and communicate that plan to others
A good communicator with excellent oral and written communication skills including evidence of preparing technical documentation and summaries for management
A natural team player able to mentor junior staff and volunteers
Desirable
Experience of delivering successful projects
Familiarity with a second language

Kew Competency Framework:

Competency	**Essential (✓)**
1.Seeing the big picture	
2.Changing and improving	?
3. Making effective decisions	✓
4. Leading and communicating	
5. Collaborating and partnering	✓
6. Building capability for all	✓
7 Achieving commercial outcomes	
8 Delivering value for money	
9 Managing a quality service	✓
10 Delivering at pace	✓