Government of **Western Australia**
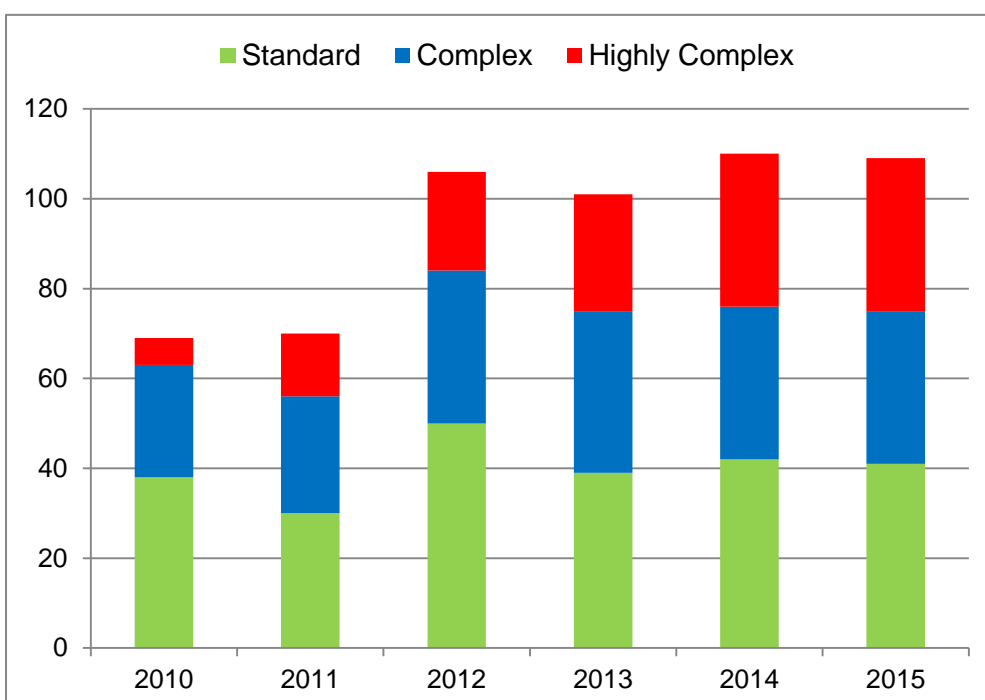Department of **Health**

Attachment 8: Complexity of Linked Data Requests Received 2010 - 2015

While the number of projects using WA linked data has increased over time, so too has project complexity. In a review of linked data requests, the Data Linkage Branch (DLB) identified five main aspects of a data request that contribute to complexity and delivery times.

New linkages	New linkages may take significant amounts of time to complete. The speed of the linkage depends greatly on the size, timeframe, quality and completeness of the dataset being linked.
Number of cohort sources	Projects become more complex as the number of cohort sources increases. The cohort selection is the first step in the preparation of a data extract for a project. Sourcing cohorts (or subsets of cases of interest) from multiple datasets takes more time. Additionally, the type and complexity selection criteria also impact on the time required.
Control Group Selections	The selection of control groups can significantly increase a project's overall complexity rating and delivery timelines. Control selection often requires a Linkage Officer and/or Data Analyst to write or amend scripts to select a suitable set of comparison records that may need to be matched, randomly selected and span multiple data sources.
Number and type of datasets	Similarly, projects become more complex and time consuming with each additional data source requested. Service data must be extracted from individual data collections for each project, so the more datasets involved, the more work required. Some datasets may have particular governance and security requirements that also impact complexity and timelines. The complexity increases for 'non-core' datasets.
Family Connections System	The Family Connections System provides genealogical links for the WA population. Projects become more complex when extracting information on family groups because it requires an additional layer of linkage and can result in multiple data extractions for each group.

Number of linked data requests 2010 - 2015

To determine the complexity 'score' of a request, each aspect listed above was assigned a number of points for certain categories:

- 2 points per new linkage required
- 1 point per cohort source
- 1 point per dataset
- 1 additional point per external "non-core" dataset
- 2 points per control group
- 5 points for use of the Family Connections system

The cut-offs for these scores were set to:

- ≤5 = Standard
- 6-10 = Complex
- ≥11 = Highly Complex