Most datasets provided externally come from sources that do not contain personal information (e.g. information about City assets). For data that contains potentially personally identifying information, appropriate mitigation strategies are developed, or data release is not recommended. Some of the mitigation strategies used to decrease the risk of personal identification include:

Anonymizing (removing) or pseudonymizing (replacing) fields containing personally identifying information. E.g. removing the owner's name from a dataset of property information.

Generalizing or aggregating data to reduce the risk of re-identifying an individual. E.g. aggregating Civic Census data to the community or ward level.

Partial release – not releasing data fields containing personal information, or suppressing aggregate records that apply to very few (<10) individuals .

These strategies help ensure data provided externally does not contain direct personal identifiers (e.g. name, phone number, identifying numbers, etc.), but can contain indirect personal identifiers (e.g. age, gender, etc.) if proper precautions are taken to mitigate the risk of individuals in the dataset being re-identified.

If the required mitigation strategies are implemented, the data may be suitable for release externally, either licensed to specific parties, or as open data. The City's Open Data Terms of Use provide an Open Government License for anyone to access and use the data for any purpose, including commercial uses, but includes exemptions that do not grant any right to use personal information.

Releasing personally identifying information for City business purposes

Approved external access to personally identifying information is rare, but sometimes required to deliver City services. When personally identifying information must be shared externally:

+ The information must only be used for the purpose for which it was collected.
+ The data is transferred using encrypted secure data transfer protocols.
+ Parties sign a binding legal agreement including restrictions on use, distribution, retention and disposition.

Future State

Early de-identification
Support efforts to de-identify data as early as possible in the data lifecycle.

Open By Design
"Open by default" addresses the question of whether data is made available externally. "Open by design" addresses the processes and systems to safely and efficiently release open data, including planning for open data at the time of planning to collect data.