Journal of Engineering and Applied Sciences 13 (4): 981-987, 2018
ISSN: 1816-949X

DataCon: Lessons Learned Enabling Easier Data Sharing, Exploration and Fusion Building a DataCon AutoGenerator Module

[1]Kyung Jin Cha and [2]Hwa Jong Kim
[1]Department of Business Administration,
[2]Department of Computer and Communication Engineering,
Kangwon National University, Chuncheon, Republic of Korea

Abstract: Data is transforming the world. Individuals, organizations, companies and governments are rushing to build technologies that generate, manage and analyze ever-increasing amounts of data. However, sharing, exploring and fusing datasets remain difficult and painful processes. We previously proposed a "DataCon" system that supports easier data sharing, exploration and fusion of many types of datasets and announced a 3 years, 1 million USD project funded by the Korean government to develop a DataCon-based data sharing platform. We now describe the lessons learned during our first phase of development: a proof of concept DataCon AutoGenerator Module which takes in arbitrary datasets and automatically generates corresponding DataCon objects. Specifically, the study describes several potential use cases for a DataCon-based data sharing platform, explores how several popular data repositories organize their datasets, sketches a preliminary data taxonomy to organize the DataCon repository, maps out a tentative technological development roadmap, recounts lessons learned implementing the initial proof of concept and lists several potential avenues for future research. We will use this study as a blueprint for future development and hope it also informs the work of others who want to make working with data easier, accelerating our collective ability to transform the world with data.

Key words: Automatic metadata generation, data fusion, metadata, tagging, summarization, DataCon

INTRODUCTION

Nowadays data has been rapidly transforming the world (Kim and Chung, 2017). Such developments enable individuals, organizations, companies and governments to promptly build technologies that generate, manage and analyze ever-increasing amounts of data (Kim *et al*., 2014; Yang, 2016). New hardware, software and algorithms are revolutionizing existing technologies and enabling establishment of new technologies in diverse domains such as transportation, logistics and e-Commerce. Although, such transformation and establishment of new technologies is most certainly beneficial, there are critical areas of concerns when it comes to sharing, exploring and fusing datasets as it remains difficult and painfully complicated processes. The primary reason for that is the lack of practical technical solutions for sharing data efficiently. This causes organizations to hoard their data in their respective silos, however there are factors that prevent successful data processes as there are legal and ethical issues like privacy concerns (Fan, 2016; Florian and Martin, 2011; Gerhard, 2002). Thus, impractical data sharing, exploration and fusion is one of the key road blocks to the achievement of a truly data-driven world (Florian and Martin, 2011; Gerhard, 2002; Venkatesan *et al*., 2017; Ugtakhbayar *et al*., 2016).

Solution; A DataCon-based data sharing platform: We previously proposed a "DataCon" system that aims to support easier data sharing, exploration and fusion of many types of datasets and announced a 3 years, 1 million USD project funded by the Korean government to develop a DataCon-based data sharing platform (Kim, 2016). This system targets provision of the practical technical solution that can be used to encourage organizations to release data beyond their silos. In addition, it also has built-in measures for legal and ethical compliance for issues like privacy management, quality assurance and other legal responsibilities.

As a concept DataCon summarizes information for the purpose of use primarily because it ensures availability and visibility of the data in the prompt and accessible manner. It can be used to assure provision of

Corresponding Author: Kyung Jin Cha, Department of Business Administration, Kangwon National University, Chuncheon, Republic of Korea

different access levels to the raw data that range from 10-100% of which depends on the needs and desires of the original data owners. Moreover, DataCon reduces the volume of traffic through minimization of redundant data and enables only required data volume. By utilizing DataCon, various data owners can choose how much of their data needs to be shared, hide private information, easily discover relevant datasets, quickly understand the important characteristics of the raw data and conveniently fuse many diverse datasets all together. It also ensures data to be visible and accessible in the prompt and accessible manner.

In further studies, we now describe the lessons learned during our first phase of development: a proof of concept DataCon AutoGenerator Module which takes in arbitrary datasets and automatically generates corresponding DataCon objects. We hope this system will be able to provide a practical solution for sharing, exploring and fusing data and accelerate the development data-driven technologies.

Use cases: There are many potential cases of use for a DataCon-based data sharing platform. Especially, it helps to solve various issues that primarily concern data owners, data users, data analysts and data engineers.

For instance, when it comes to data owners, a DataCon-based data sharing platform enables them to promote their data to data users. At the same time, data users can download DataCon samples in order to decide whether it is worthy to obtain the full dataset of interest. Thus, it gives them an awareness and understanding to make wiser data acquisition decision. Moreover, with DataCon data analysts are enabled to find partners to perform mutually beneficial data fusion. Although, concept of the data fusion is not novel and has been considered as increasingly feasible, it has been confronted by many challenges that prevent successful accomplishment of the data fusion techniques. Such include risks of being potentially involved in the issues of privacy or legal matters, complications that relate to guarantee future presence and quality of data; lack of feedback regarding usefulness of the released data, lack of monetary compensation in the case of the free data release and lastly, lack of the convenient methods to share data partially. Thus, use of the DataCon-based data sharing platform can potentially minimize these risks and contribute to the successful data fusion. Furthermore, the utilization of DataCon can enable data engineers to make better decisions regarding how to prepare the data. As more companies nowadays understand that it is imperative to become data-driven to maintain competitiveness, data preparation becomes crucial. However, it typically requires data professionals to spend a lot of time and effort to prepare data which takes time away from data analytics itself. Such issues are caused by the versatility of datasets and overall lack of the established standardized rules as most of the data professionals use their own methodologies in data preparation processes which are not collected nor accessible by others. Thus, use of DataCon-based data sharing platform enables companies to make right decisions data preparation processes and potentially speed it up which in turn enables greater competitiveness as well as competence.

The automatically generated attributes of the DataCon-based data sharing platform can enable analysis across multiple sources of data and multiple types of data. For instance, DataCon enables analyzing multiple CCTV data feeds simultaneously to find a person moving across the cameras and checking nearby public transportation and event schedules to guess why the person is moving in that direction.

In addition, one of the most critical aspects of the valuable data-driven applications often regards personalization. Nevertheless, personalization by definition involves sensitive data which concerns privacy management. With the utilization of DataCon-based data sharing platform sharing and exploration of such sensitive datasets will be enabled to safeguard the privacy of individuals. Privacy management concerns have always been one the most critical issues. From the very beginning utilization of the DataCon enables automatization of the data preparation process which allows to encode privacy safeguards. Such encoding regards person that performs the analysis as well as who will be able to see the results.

DataCon AutoGenerator Module: The DataCon AutoGenerator Module takes in arbitrary datasets and automatically generates corresponding DataCon objects. The DataCon objects are intended to make it processes of dataset's sharing, exploration and fusion prompt and easier. This is done by automatically summarizing the key characteristics of each datasets to accelerate exploratory analysis. In the future, this structured summary information can be used to recommend useful datasets for your particular task as well as automate data preparation and data fusion for later analysis. The primary purpose of this study was to start exploring what those key characteristics might look like for different types of datasets.

All DataCons share a common set of attributes like size, date, owner and percentage of raw data to be shared. Additionally, each DataCon will have additional attributes

based on the specific type of the dataset. For example, an image dataset may contain automatically generated captions while a text dataset may contain bag of words counts.

This study explore what those common attributes and type-specific attributes look like. We have implemented a DataCon AutoGenerator Module using Python 3 that takes in arbitrary datasets and outputs a DataCon object with many of these attributes. We will continue to develop this module over time by refining how we categorize data and implementing additional type-specific attributes.

MATERIALS AND METHODS

We explored the following popular data repositories to see how they organized their datasets:

- UCI machine learning repository (Anonymous, 2000)
- Data.gov (Anonymous, 2006)
- The World Bank Data (Anonymous, 2013)
- Data.go.kr (2016a, b)
  Kaggle Datasets (Anonymous, 2018a)
- Census.gov (Anonymous, 2018b)
- AWS public datasets (Anonymous, 2018c)
- Google public datasets (Anonymous, 2018d)

The repositories were chosen because they are some of the most commonly used repositories for data science projects in Korea, America and the rest of the world. We came across an interesting insight that each repository organized its data in a different way, although there is an overlap in several of their fields.

Common DataCon attributes

Name: Every dataset has a human-generated or computer-generated and human-approved name, often representative of the topic, provider or location of the data. Examples include "Hacker News", "IRS 990 Data", "MNIST" and "World Bank: Educational Statistics".

Datasets are often presented in a list ordered alphabetically by name. For example, Google public datasets lists 38 datasets sortable only by name. However, when you don't already know the name of the dataset you're searching for youre limited to searching for keywords and hoping for hits, randomly scrolling through the list or exhaustively checking each of the many datasets one-by-one. When there are too many datasets, this quickly becomes hugely inefficient.

However, once the number of datasets has been narrowed down, alphabetic order by name is often the default way to display lists of datasets.

Address: These names are useful for humans to refer to the datasets but computers need something more explicit such as a file path once the datasets are downloaded or IP addresses in the form of a URL an API or a server address that, combined with the appropriate credentials, lets people download datasets.

As these addresses are meant for computer consumption, not human consumption, people rarely if ever sort datasets by address. Instead, they hide the addresses in the technical documentation, hyperlinks or software files.

Topic: The most common way of sorting datasets is performed by topic. Here are some of the topics that chosen repositories use: UCI Machine Learning Repository (Anonymous, 2013) Life Sciences, Physical Sciences, CS/Engineering, Social Sciences, Business, Game, Other. Data.gov (Anonymous, 2016a, b) Agriculture, Climate, Consumer, Ecosystems, Education, Energy, Finance, Health, Local Government, Manufacturing, Maritime, Ocean, Public Safety, Science and Research. The World Bank Data (Anonymous, 2013) Agriculture and Rural Development, Aid Effectiveness, Climate Change, Economy and Growth, Education, Energy and Mining, Environment, External Debt, Financial Sector, Gender, Health, Infrastructure, Poverty, Private Sector, Public Sector, Science and Technology, Social Development, Social Protection and Labor, Trade, Urban Development. Data.go.kr (Anonymous, 2018a) Education, Land, Administration, Finance, Industry and Employment, Welfare, Food and Health, Culture and Tour, Medicine, Safety, Transportation, Environment and Weather, Science and Research, Agriculture, Diplomacy, Law. Census.gov (Anonymous, 2018b) Population, Economy, Business, Education, Emergency Preparedness, Employment, Families amd Living Arrangements, Health, Housing, Income and Poverty, International Trade, Public Sector. AWS Public Datasets (Anonymous, 2018c) Astronomy, Biology, Chemistry, Climate, Economics, Encyclopedic, Geographic, Mathematics.

It is important to note that each of these topics can have subtopics and subsubtopics and so on for however many layers makes sense for the datasets at hand. None of these repositories use the exact same list of topics. This implies that the best topics to sort by depend on the datasets and the users you expect to have on your repository.

Source: Datasets can also be sorted by the person, sensor or organization that produced the data. For example, Census.gov allows you to sort datasets by the survey that was used to generate the data (Anonymous,

2018c) and Data.gov lets you sort by Organization Type, Organizations and Bureaus (Anonymous, 2018a).

Publisher: The person who uploaded the data to the repository is often not the person who generated the data. This means many repositories distinguish uploader and source. On Kaggle Datasets (Anonymous, 2018c) for example, you can go to specific user profiles and see all the datasets they've uploaded and Data.gov lets you search by Publisher (Anonymous, 2018a).

Geographic region: Some datasets are tied to specific locations and can be sorted by continent, hemisphere, country, region or locality. For example, The World Bank Data lets you sort datasets by country (Anonymous, 2018d).

Format: Each dataset is stored in a specific format. Some repositories allow you to search by format. For example, Data.gov lets you search by each of the following formats (the number in parenthesis after the format type represents how many datasets in the particular format are present in the repository) (Anonymous, 2018a):

> HTML (76118), PDF (41764), XML (33939), Originator data format (26145), ZIP (21203), CSV (16474), TIFF (13653), JSON (13140), MrSID (12890), text/xml (12806), WMS (11710), RDF (9156), XYZ (7960), JPG (7568), Esri REST (6638), WCS (5624), TXT (4928), NetCDF (4440), KML (4235), application/octet-s... (3642), IWXMM-US (3560), WFS (2706), Excel (2701), gml (2362), tif (1897), HDF (1749), ESRI Shapefile (1574), EXE (1391), JPEG (1301), GeoJSON (1268), XLS (1262), OPeNDAP (957), FEMA-DCS-Hydrology (949), XLSX (924), FEMA-DCS-Hydraulics (918), ARCE (905), TAR (849), None (804), GeoTIFF (803), FEMA-DCS-Terrain (792), ASCII (675), application/vnd.lot... (613), application/vnd.goo... (602), application/vnd.goo... (597), SHP (574), KMZ (563), API (509), Original metadata r... (505), OGC WMS (429)

In addition, it's also possible to sort by popular formats only or by general binary decisions like "geospatial or non-geospatial" and "matrix or non-matrix".

Data type: Every dataset is a collection of data and often one or a few types of data, for example "numerical" and "categorical". Many repositories allow you to sort by data type. The UCI machine learning repository, for example, sorts by multivariate, univariate, sequential, time-series, text, domain-theory and other (Anonymous, 2000).

Newest: Many repositories show data sorted by "Newest" which can mean recently uploaded, recently updated, recently viewed or all of the above. The idea is that newer datasets are more likely to be interesting because of their novelty. For example, the UCI machine learning repository has a newest list of datasets sorted by most recent date of upload first (Anonymous, 2000).

Most popular: Many repositories show data sorted by "Most popular" which is often defined by most upvotes, most views, most comments or some combination of many factors. The idea is that datasets that are interesting to many people are more likely to be interesting to you, too. For example, the UCI Machine Learning Repository has a Most popular list of datasets sorted by number of hits, since, 2007.

Task: Some datasets are generated for a specific purpose. For example, the Abalone dataset on the UCI Machine Learning Repository was generated so people could use data analysis algorithms to predict the age of abalones from physical measurements, originally structured as a classification task (Anonymous, 2008).

However, not all datasets were generated with a specific task in mind. Additionally, many datasets, even datasets that were generated with a specific task in mind, can be used for many tasks: classification of many different target variables, regression, clustering and visualization. Instead of specifying a single task for each dataset, it may make more sense to allow users to add as many "possible tasks" as desired, provided each task label was accompanied by a short description of how the variables of that dataset could be used to accomplish that task.

Featured: Many repositories, like the UCI machine learning repository, allow manually chosen datasets to be featured on the front page of the repository, often accompanied by short descriptions of why that dataset was chosen (Anonymous, 2000).

Number of rows, columns: Many datasets are in a matrix format and thus have a specific number of rows or a specific number of columns. This makes it possible as on the UCI machine learning repository to sort by the number of rows or columns (or both) of the dataset (Anonymous, 2000). Many datasets, however, are not in matrix format and thus need other metrics like "number of records" or "number of objects" to accomplish the same goal.

Size: Another metric that can be used to characterize datasets is size: the number of bytes it will take up when downloaded. Searching by size can help you limit your search to datasets that will fit on your available memory. However as it's possible to downsample larger datasets or easily purchase more computing power on cloud

computing services, the size of a dataset may be best used to give the user a feel for the computing requirements of working with that dataset.

Usage rights: Some repositories allow you to search by the usage rights or license that the data was released under. For example, Data.go.kr lets you search by Copyright indication, the copyright indication-modification prohibition, the copyright indication-same condition border permission, the copyright indication-non-profit, the copyright indication-non-profit-modification prohibition, the copyright indication-non-profit-same condition border permission and the use permission range limitless.

Percentage of data shared: Many datasets that are publically released are released in their entirety. Some datasets, however are only partially released to advertise that the dataset exists and encourage people who want the full dataset to contact the data publisher.

Tags: The most flexible data repository sorting system involves repository- or user-defined tags. These tags can be anything, including anything in the aforementioned categories. Repositories may incorporate tags so that users can incorporate information they think is important but is not included in the system-defined labels. However, because users define tags, there may quickly become a large number of nonsensical or redundant tags and other users may not know what tags to search for.

Common DataCon attributes: The DataCon platform should use all of the above methods and allow users to organize, search and sort DataCons by:

- Name
- Address
- Topic
- Source
- Publisher
- Geographic region
- Format
- Data type
- Date uploaded
- Date updated
- Views
- Task
- Featured
- Number of rows, columns (or equivalent for non-matrix datasets)
- Size
- Usage rights
- Percentage of data shared
- Tags

The DataCon AutoGenerator Module should thus automatically generate each of these characteristics for arbitrary datasets.

RESULTS AND DISCUSSION

Type-specific datacon attributes: All DataCons share a common set of attributes like size, date and owner as well as additional attributes based on the specific type of the dataset. The previous section covered attributes common amongst all DataCons. This section explains how we categorized different types of data and our list of type-specific DataCon attributes for each data type.

Types of data: There are many ways to taxonomize data type. An exhaustive taxonomy is perhaps impossible as there are many subtypes depending on how you define data: you could have a type of data called neuroscience data with subtypes called EEG, MEG, fMRI and CAT (and the many other neuroscience sensors) each with sub-sub types depending on the specific machine and mode used. Repeat this for every scientific, engineering and technical discipline and you'll find that generating a truly accurate data type taxonomy would require a massive effort with data users from every field and many countries around the world.

For DataCon we start with the following categories: "Numerical", "Categorical", "Image", "Text" and "Mixed and Other". As more datasets are seen in the DataCon repository, more categories could be added. We have implemented many attributes for each type of data. For instance, for numerical data type additional attributes could be median, quartile values, outliers, missing values, principal components, etc. For text type data, in addition to the current attributes of corpus, bi-to-quadgram TFIDF values and trigram term frequencies, attributes of bag of word counts, the title of the document, the entities in the document, grammatical structure, text summary, clusters of similar documents, word cloud, etc., could be potentially added. All of the attributes that could be added and implemented in the future for each data category are listed and displayed in Table 1.

Although, we were able to arrive at the categorization and describe its processes of different types of data, list type-specific DataCon attributes for each data type as well as list potential attributes catering to category-specific data that could be implemented in the future, there are several avenues for future research on a DataCon-based data sharing platform. The first is

Table 1: Datacon type-specific attributes, current and future

Type of data	Currently implemented attributes	Potential future attributes
Numerical	Values is_null, num_null, maximum, minimum, mean, range is_outlier, values_sans_outliers, boxplots and histograms	Median, quartile values, outliers, missing values, principal components and additional visualizations
Categorical	Values is_null and num_null,and bar plots	Additional visualizations
Image	The mean image as well as for each image, the pixel values, the histogram of oriented gradients pixel values and visualizations of the images	Automatically generated captions, clusters of similar images, shapes and boundaries between different image regions
Text	For the corpus, bi-to-quadgram TFIDF values and trigram term frequencies	Bag of word counts, the title of the document, the entities in the document, grammatical structure, text summary, clusters of similar documents, word cloud, language of document
Mixed and other	Currently, we do not have additional attributes for mixed and other types of data. As, we see more datasets and refine our data taxonomy further we will continue adding structured categories and defining lists of attributes for them	Potential future types of data include video time series and geospatial data

extending work on the DataCon AutoGenerator Module. The second is prototyping the data sharing platform. The third is exploring additional use cases for the DataCon object.

DataCon AutoGenerator Module: Additional work on the DataCon AutoGenerator Module involves adding additional automatically generated attributes and refining the DataCon data taxonomy. We have implemented many common and type-specific attributes and listed many additional future type-specific attributes in Table 1. However, to make sure these attributes are useful we need to seek out feedback from various data users. This involves discussing our proposed solution with data owners, analysts, engineers, scientists and other data users. Additionally, we need to make sure our system complies with international metadata standards.

Currently, the DataCon AutoGenerator Module is a custom Python 3 module. However, to make generating DataCon objects easy we need to develop a web interface that allows users to upload their datasets (or connect to their databases) and outputs a DataCon object. Additionally, the DataCon object is currently a custom Python object. This means saving and loading the DataCon objects requires using Python. We need to develop a custom file format and loading software that non-coders can use.

Data sharing platform: When the DataCon objects are created, they need to be easily discoverable. This involves creating a web application that allows users to organize, search and sort DataCon objects by the common DataCon attributes. This web application should understand the user's goal and search patterns and recommend potentially useful DataCons. Additionally, this web application needs to allow users to easily interface with the DataCon AutoGenerator Module to create DataCons for their own datasets as well as download existing DataCons. Users should also not have to download DataCons to explore them. The web application should give users a graphical user interface that allows them to explore the attributes of individual DataCon objects. This interface should also allow for Jupyter notebook-style programmatic interactions. Finally, the DataCon platform needs to have built-in measures to comply with legal and ethical requirements on issues like data privacy.

Additional datacon uses: Currently, DataCon objects are intended to facilitate easier data sharing, exploration and fusion. However, in the future, DataCons may be used to automate data preparation and analysis.

CONCLUSION

Companies, organizations and governments are building technologies that generate, manage and analyze ever-increasing amounts of data. However, sharing, exploring and fusing datasets remain difficult and painful processes. The proposed "DataCon" system supports easier data sharing, exploration and fusion of many types of datasets. In this study, we described our lessons learned from developing a proof of concept Datacon AutoGenerator Module which takes in arbitrary datasets and automatically generates corresponding DataCon objects. We will continue to develop this system in hopes of providing a practical technical solution that can be used to encourage organizations to release data beyond their silos and also has built-in measures for legal and ethical compliance for issues like privacy. We believe sharing our efforts will help inform the work of others who want to make working with data easier and accelerate our collective ability to transform the world with data.

ACKNOLEDGEMENTS

This research was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) through (Advanced Production Technology Development Program) funded by Ministry of Agriculture, Food and rural Affairs (MAFRA) (116116-03-1-SB010) and by Institute for Information and communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No: R-20160906-004163, Developing Bigdata Autotagging and Tag-based DaaS System)

REFERENCES

Anonymous, 2000. Abalone data set. National Science Foundation, Virginia, USA. http://archive.ics.uci.edu/ml/datasets/Abalone.

Anonymous, 2006. Google big query public datasets. Wix.com Web Development Company, Tel Aviv israel. https://cloud.google.com/bigquery/public-data/.

Anonymous, 2013. UC irvine machine learning repository. National Science Foundation, Virginia, USA. http://archive.ics.uci.edu/ml/index.php.

Anonymous, 2016a. AWS public datasets. Amazon Web Services, Inc., Seattle, Washington, USA. https://aws.amazon.com/datasets.

Anonymous, 2016b. Access data through products and tools including data visualizations, mobile apps, interactive web apps and other software. United States Department of Commerce, Washington, DC., USA. https://www.census.gov/data.html.

Anonymous, 2018a. Open data day in Korea. Ministry of the Interior and Safety, Seoul, South Korea. https://www.data.go.kr/main.do?lang=en.

Anonymous, 2018b. The home of the US Government's open data. USA. https://www.data.gov/.

Anonymous, 2018c. Welcome to Kaggle Datasets the best place to discover and seamlessly analyze open data. Kaggle, San Francisco, California, USA. https://www.kaggle.com/datasets.

Anonymous, 2018d. World Bank open data. World Bank, Washington, DC., USA. https://data.worldbank.org/.

Fan, P., 2016. Coping with the big data: Convergence of communications, computing and storage. China Commun., 13: 203-207.

Florian, B. and K. Martin, 2011. Linked Open Data: The Essentials. Druck & Grafik, Vienna, Austria,.

Gerhard, S., 2002. Data privacy approaches from US and EU perspectives. Telematics Inf., 19: 193-200.

Kim, G.H., S. Trimi and J.H. Chung, 2014. Big-data applications in the government sector. Commun. ACM., 57: 78-85.

Kim, H.J., 2016. DataCon: Easier Data Sharing, Exploration and Fusion with Automatic Metadata Generation. In: AI 2016: Advances in Artificial Intelligence, Kang, B. and Q. Bai (Eds.). Springer, Switzerland isBN:978-3-319-50126-0, pp: 708-713.

Kim, S. and K. Chung, 2017. Method of recovery of deleted records in a postgre SQL database. Intl. J. Technol. Eng. Stud., 3: 169-176.

Ugtakhbayar, N., B. Usukhbayar, S.H. Sodbileg and J. Nyamjav, 2016. Detecting TCP based attacks using data mining algorithms. Intl. J. Technol. Eng. Stud., 2: 1-4.

Venkatesan, N.J., C.S. Nam, E. Kim and D.R. Shin, 2017. Analysis of real-time data with spark streaming. J. Adv. Technol. Eng. Res., 3: 108-116.

Yang, F.J., 2016. The user interface design of an intelligent tutoring system for relational database schema normalization. Intl. J. Technol. Eng. Stud., 2: 70-75.