Data communication user research in the Post Office: Part 2

R. Jorré*, G. Boyfield†, and J. Darkin†

Part 1 of this paper described in general terms the approaches that have been made, that are in progress, and that are proposed for research by the Post Office into customer use of, and demand for, data transmission services. Part 2 presents a set of carefully defined categories of information from which, it is believed, models of EDP usage and subsequent development of Data Transmission (DT) could be constructed.

(Received June 1973)

1. Introduction

Part 1 of this paper has described in general terms the approaches that have been made, that are in progress, and that are proposed for research by the Post Office into customer use of, and demand for, data transmission services. It is believed that the future demand for data transmission is closely connected not only with its present use, but also with the usage of, and experience with computers in the United Kingdom. Part 2 presents a set of carefully defined categories of information from which, it is believed, models of EDP usage and subsequent development of Data Transmission (DT) could be constructed. These concepts are considered to be of wide general interest to readers.

Motivations

There are many possible motivations for extending the efficiency of computer operation by the incorporation of data transmission facilities. The motivations include, for large organisations economies of scale in computer use, which may be in specialised computer personnel, computer storage, hardware, or the development of specialised software. These advantages could be realised by the replacement of many small computer installations scattered in establishments throughout the country by one or two large computers placed in centralised locations.

The possibility of creating large centralised files is another aspect of the economies of scale which enables organisational operations such as centralised bulk buying to be carried out much more effectively and efficiently than in the past. It also facilitates new types of analyses on operational performance. The accuracy of data input can be substantially increased by referring back to the point of contact with the outside world rather than data preparation being carried out through a series of intermediaries.

The potential for increased speed of organisational response to outside stimuli made possible by data transmission, allows for inventory reduction and better resource planning.

The provision of inter-active computing enabling a dialogue to take place between a computer user and a computer may also lead to considerable improvement in programming and computer efficiency.

Whilst many of the motivations described above apply only to large organisations it is clear that small organisations also benefit by the resource sharing that a combination of the computer bureau industry and data transmission has made possible. This sharing not only applies to hardware or machine time but also enables small organisations to gain easy access to specialised software packages from geographically dispersed bureau services.

It is clear from the foregoing description of possible motives for data transmission demand that its nature may vary widely from industrial sector to industrial sector, that it may be critically dependent on whether an organisation within a particular sector is large or small, whether its establishments are widely dispersed or in close geographical proximity. It is also likely to depend upon the experience of computer use which exists within the sector and the amenability of the sector's problems to computer solution. Also account must be taken of the financial success of the industrial sector concerned and its likely growth rate.

A model of DT demand

Imminent users of data transmission, those for whom the above motivations are the strongest are, it is believed, users with considerable experience of computers without data transmission and are in industrial sectors where there is already considerable reliance on computing methods. The theoretical advantages of producing a model for data traffic forecasting have already been discussed in Part 1. It remains therefore to define a suitable model which could take into account the driving forces towards data transmission and then, again with the model in mind, to define the data that is necessary for its implementation. With the data defined, methods of collection, storage and analysis can be prepared.

The model described in the first section below is essentially a probabilistic formulation to describe the progress of an organisation or, more generally 'user organism', categorised by type, sector, or other classification found by subsequent research to be important, as it develops, in sequence, a number of EDP applications. With each development a probability of data transmission can be associated. It should be mentioned here that the actual mathematical detail of the second part of the modelling section can be passed over by those readers only concerned with basic concepts.

Definitions

The decision to model computer use in organisations and to define this usage in terms of the historic adoption of applications leads directly to the problems of the definition of EDP applications. This subject is described fully in the second section. It is sufficient here to state that the definition chosen must be a quantifiable entity with clearly defined measures of its size and the impact that it has on the users' organisational structure. It is believed that the lack until now of such an application definition has proved not only to be an impediment to the forecasting of data transmission demand but also a blockage to forecasting computer usage in general. For this reason it is hoped that the general techniques proposed for tackling a problem of forecasting data transmission demand

*Statistics and Business Research Department, The Post Office, 23 Howland Street, London W1P 6HQ,
†Telecommunications Marketing Department, The Post Office.

are helpful to those who wish to analyse more fully the impact of computers upon our industrial society.

In addition to defining application, the model demands that strict definitions are given to the parameters of the organisation and, indeed, to the definition of organisation itself. It is clear from the variety of possible motivations to data transmission that the organisation must be described by many parameters. The third section, on terminology, deals with this important subject of general definitions.

Data collection

Section 4 outlines the gathering of the information and the specific items of data involved. As the information is large in volume and comparatively complex, it may be necessary to obtain it from a number of respondents in any particular organisation, each one being able to contribute only a portion of the total required information. Also plans must be made for this information to be up-dated at comparatively frequent intervals so as to enable an historic picture to be built up of the flows of applications through organisations and industry sectors.

A model of data transmission demand

This section describes a model which, it is believed, represents the process taking place within a user organism as it passes through the various stages of electronic data processing (EDP) sophistication. Furthermore the model allows the incorporation of general overall relationships which may exist between the user organism and its environment.

The representation described is an example of the type of approach which could prove relevant in analysing and forecasting data transmission (DT) usage although, obviously, it will almost certainly have to be considerably modified as more detailed information becomes available. It is important, however, in focusing attention on such points as the definition of applications, user organisms and EDP equipment and by showing data requirements. It is also an initial link in a chain of model building, data collection and consequent model modification by which, it is hoped, present rather subjective methods may be replaced by more objective research.

An effort has been made to avoid a 'black box' model as these usually can do no more than describe a system by a series of convenient mathematical equations involving parameters which are difficult or impossible to interpret. The present situation is more exigent. Model formulation must be directly linked to the causal mechanism, this is a necessity if it is to be used for predicting the behaviour of the system when subject to different constraints or stimuli. This type of manipulation is usually done by changing parameter values, obviously only feasible when these have a definite real-world meaning.

Description

The flows in the model for any one and only one user organism are illustrated in **Fig. 1.** Three sets of elements are involved:

1 *Set of possible EDP applications*
This is the number of EDP applications which have been technically perfected or found feasible and are relevant to the particular industrial sector or other useful grouping of which the user organism is a part but which are not applied by it. The content of this set is represented as a random variable dependent on economic factors related to the particular industry or some other useful grouping.

2 *Set of EDP applications in use without DT*
The user organism draws from the set of possible EDP applications and applies a certain number of them. These, until they are linked with data transmission, form the second set of elements. Content is a random variable dependent on business, social and economic variables related to the user organism itself.

3 *Set of EDP applications with data transmission*
The user organism can draw from the set of possible EDP applications to form this third set of applications with data

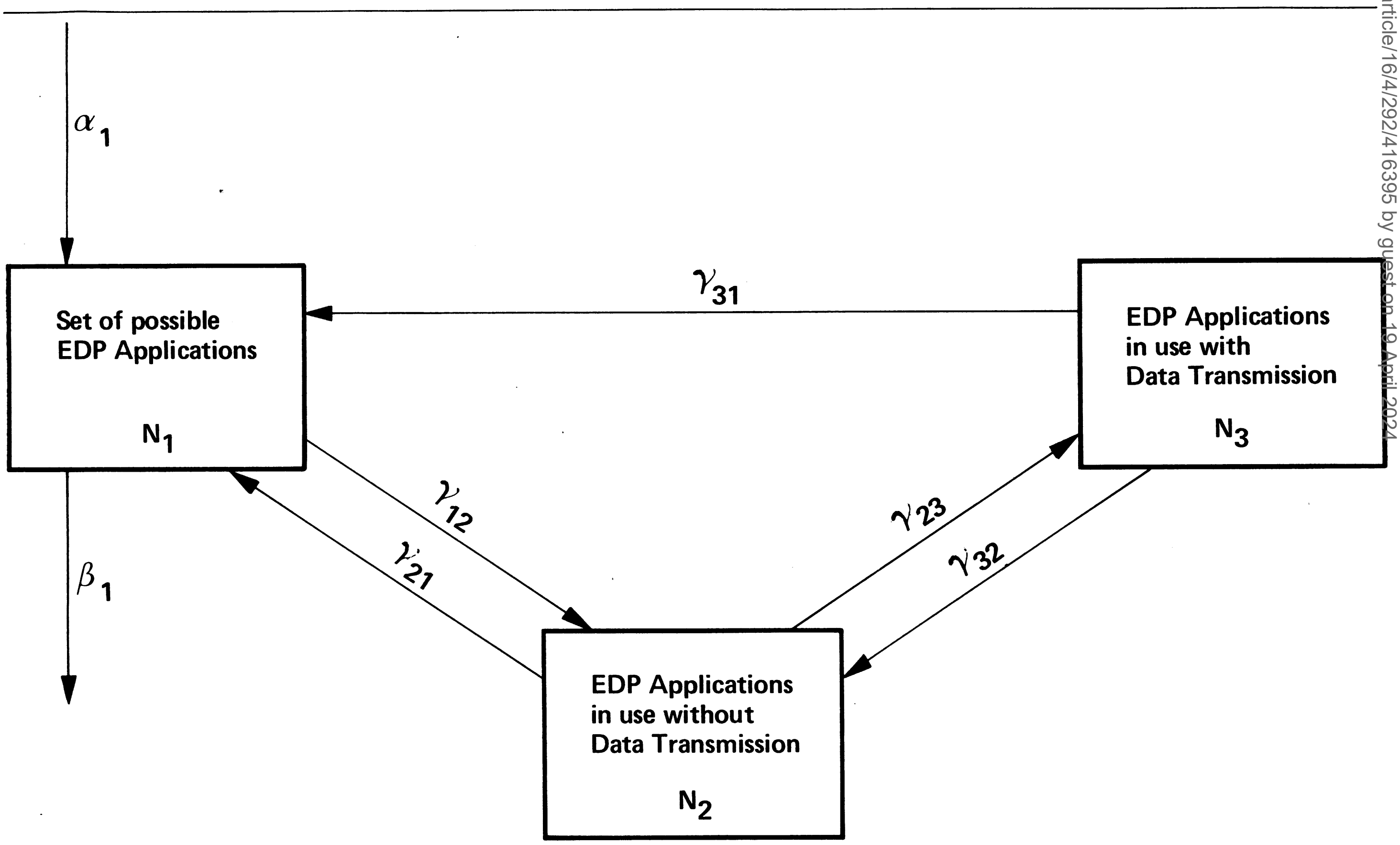

Fig. 1 A general model for flows of applications in any one user organism

transmission. Content is once again a random variable linked to the business, social and economic characteristics of the user organism. These latter may or may not be the same as those affecting the previous set and the functional relationship may or may not be similar.

The contents of the sets, therefore, are dependent on economic, organisational and other business characteristics which have been found relevant to the phenomena of computer applications and data transmission usage. As a further generalisation, it is also possible to link the flow of elements through the sets at any time to the content of the originating set.

Seven possible flows have been postulated in the model and are described below:

(*a*) *New inventions and feasibility studies*
New inventions enter the set of possible EDP applications.

(*b*) *Discarded inventions*
When an invention is no longer found useful it leaves the set of possible EDP applications.

(*c*) *EDP usage*
When the user organism decides to apply a new EDP application it effects a transfer from the set of possible EDP applications to the set of those in usage.

(*d*) *EDP discard*
If the user organism after a certain time finds an EDP application no longer needed or desired for organisational, financial or other reasons it effects a transfer from the set of applications in usage to the set of possible applications. Eventually such an application can completely leave the system by the 'discarded inventions' flow.

(*e*) *DT usage*
Once the user organism is using the application it can decide to refine it by incorporating data transmission. It then effects a transfer from the set of EDP applications in usage to the set of DT applications in usage.

(*f* and *g*) *DT discard*
If the user organism after a certain time finds that the incorporation of DT is no longer wanted it can effect a transfer back to the set of EDP applications in usage. If however the whole application is found to be inadequate at this stage, it can effect a transfer back to the set of possible EDP applications.

Notation and specification

Consider a set θ of user organisms having a particular member θ_0. Then:

N_θ: Number of user organisms in θ

$\alpha_1(\theta, t)$: Instantaneous rate of incorporation into set 1, that of 'possible EDP applications'. This parameter is a function of relevant economic variables linked to the set θ and time t.

$\beta_1(\theta, t)$: Instantaneous rate of discard from set 1. Also a function of θ-variables and t.

$N_1(t)$: Number of applications in set 1 at time t.

$N_2(t)$: Number of applications in set 2, that of 'EDP applications without DT' at time t.

$N_3(t)$: Number of applications in set 3, that of 'EDP applications with DT' at time t.

$\gamma_{ij}(\theta_0, t)$: Instantaneous rate of transfer from set i to set j. These are functions of economic variables linked to the individual user organism, θ_0, and t.

$\phi_j(N_j)$: Functions linked to the number of elements in set j. The state of θ_0 at time t is given by $N_1(0) = i_0$, $N_2(0) = j_0$, $N_3(0) = k_0$.

Throughout the rest of this section it is assumed that t is continuous in the intervals considered.

The reason for formulating continuous time rather than the discrete variant, more in line with the data, is that it is easier to obtain an analytic solution with the former and this is an approximation to the characteristics of the discrete imbedded process (Cox and Miller, 1965).

It is assumed that the seven movements shown in Fig. 1 are independent Poisson processes specified as follows:

$$Pr\{N_1(t, t + \Delta t) = 1, N_2(t, t + \Delta t) = 0, N_3(t, t + \Delta t) = 0\} = \alpha_1(\theta, t)\,\Delta t + o(\Delta t) \quad (1)$$

$$Pr\{N_1(t, t + \Delta t) = -1, N_2(t, t + \Delta t) = 0, N_3(t, t + \Delta t) = 0\} = \beta_1(\theta, t)\,\Delta t + o(\Delta t) \quad (2)$$

$$Pr\{N_1(t, t + \Delta t) = 1, N_2(t, t + \Delta t) = -1, N_3(t, t + \Delta t) = 0\} = \gamma_{21}(\theta_0, t)\,\Delta t + o(\Delta t) \quad (3)$$

and similarly for the remaining γ_{ij}

The probability of 'no movement' results from a compounding of the seven movements described above.

$$Pr\{N_1(t, t + \Delta t) = 0, N_2(t, t + \Delta t) = 0, N_3(t, t + \Delta t) = 0)\} = 1 - S\Delta t + o(\Delta t) \quad (4)$$

where

$$S = \alpha_1(\theta, t) + \beta_1(\theta, t) + \gamma_{12}(\theta_0, t) + \gamma_{21}(\theta_0, t) + \gamma_{32}(\theta_0, t) + \gamma_{23}(\theta_0, t) + \gamma_{31}(\theta_0, t)$$

If probability intensities α_1, β_1, γ_{ij} change according to the number of elements in the set from which movement takes place, then this modification can be of the form:

$$\begin{matrix} \alpha_1 \\ \beta_1\,\phi_1\,(N_1) \\ \gamma_{ij}\,\phi_i\,(N_i) \end{matrix} \quad (5)$$

where $\alpha_1(\theta, t)$ has been abbreviated to α_1, $\beta_1(\theta, t)$ to β_1, $\gamma_{ij}(\theta_0, t)$ to γ_{ij} and $N_i(t)$ to N_i.

Then under these conditions the process is completely specified by the usual recurrence relationship and boundary condition.

Simplification and solutions

It is useful to specify the various parameters involved in the model in a fairly simple way as this facilitates the setting up of null hypotheses against which to test real data. Analytic rather than simulated results are more efficient in this type of situation and are also useful for establishing conditions under which the model will hold.

It is assumed that:

(*a*) All parameters are time homogeneous, thus $\alpha_1(\theta, t) = \alpha_1(\theta)$, $\beta_1(\theta, t) = \beta_1(\theta)$, $\gamma_{ij}(\theta, t) = \gamma_{ij}(\theta)$ $\forall i, j$ and these are all abbreviated as in 5.

(*b*) $\phi_i(N_i) = 1$ for all $N_i > 0, \forall i$
$= 0 \quad N_i = 0, \forall i$.

Movement is therefore thought as independent of the number of elements in the set from which it takes place. This is not really necessary for equilibrium solutions but becomes desirable for finite t.

(*c*) User organisms in a set θ are thought to act independently.

1. *Equilibrium solution*

The equilibrium solution for this model has been obtained in explicit form (Kingman, 1969; Whittle, 1967; Whittle, 1968). For any one user organism it is given by:

$$Pr\{N_1 = n_1, N_2 = n_2, N_3 = n_3\} = \text{Const.} \prod_{s=1}^{s=3} \left[\frac{\lambda_s^{n_s}}{\prod_{i=1}^{i=n_s} \phi_s(i)} \right] \quad (6)$$

where $s = 1, 2, 3$ referring to the sets and

$$\lambda_1 = \frac{\alpha_1}{\beta_1}. \quad (7)$$

$$\lambda_2 = \frac{\alpha_1 \gamma_{12}(\gamma_{31} + \gamma_{32})}{\beta_1(\gamma_{21}\gamma_{31} + \gamma_{21}\gamma_{32} + \gamma_{23}\gamma_{31})}. \quad (8)$$

$$\lambda_3 = \frac{\alpha_1(\gamma_{12}\gamma_{23})}{\beta_1(\gamma_{21}\gamma_{31} + \gamma_{21}\gamma_{32} + \gamma_{23}\gamma_{31})}. \quad (9)$$

Furthermore, due to simplification:

$$Pr\{N_s = n_s\} = (1 - \lambda_s)\lambda_s^{\,n_s} \quad n_s = 0, 1, 2 \quad |\lambda_s| < 1 \quad (10)$$

If the probability generating function of N_s is $E_s(z)$, then

$$E_s(z) = \sum_{n_s=0}^{n_s=\infty} Pr\{N_s = n_s\} z^{n_s}.$$

$$= \frac{(1 - \lambda_s)}{1 - \lambda_s z} \quad (11)$$

Now on generalisation to the N_θ user organisms two distinct cases arise:

(*a*) $s = 1$
Here λ_1 depends only on α_1, β_1 which in their turn are functions of θ. The central limit theorem applies as $N_\theta \to \infty$ or, for small N_θ a specialisation of (*b*) below, is possible

(*b*) $s = 2, 3$
In this case λ_s will vary from user organism to user organism dependent on individual values of θ_0.
Now if $\lambda_s^{(i)}$ is the parameter for the *i*th user organism $i = 1, 2, 3 \ldots N_\theta$, then because of simplification (*c*), the generating function of the total numbers in set *s*, summed over all *i*, $E_s^\theta(z)$, is given by

$$E_s^\theta(z) = \prod_{i=1}^{i=N_\theta}\left(\frac{1 - \lambda_s^{(i)}}{1 - \lambda_s^{(i)} z}\right) \quad (12)$$

$$= \left[\prod_{i=1}^{i=N_\theta}(1 - \lambda_s^{(i)})\right]\left[\exp\left\{\sum_{r=1}^{r=\infty}\left[\sum_{i=1}^{i=N_\theta}(\lambda_s^{(i)})^r\right] z^r \lambda^{-1}\right\}\right] \quad (13)$$

Individual probabilities, means and variances can now be extracted in the usual manner.
This process is considerably facilitated by writing:

$$E_s^\theta(z) = \sum_{m_s=0}^{m_s=\infty} Pr\{M_\theta^s = m_s\} z^{m_s}$$

where M_θ^s is the total in set *s* across all $\theta_0 \in \theta$.
Then

$$E_s^\theta(z) = \sum_{m_s=0}^{m_s=\infty} m_s!\, Pr\{M_s^\theta = m_s\} z^{m_s}/m_s!$$

$$= \sum_{m_s=0}^{m_s=\infty} g_{m_s} z^{m_s}/m_s!$$

which by (13)

$$= \left[\prod_{i=1}^{i=N_\theta}(1 - \lambda_s^{(i)})\right]\left[\exp\left\{\sum_{r=1}^{r=\infty} A_r z^r/r!\right\}\right]$$

$$A_r = \sum_{i=1}^{i=N_\theta}(\lambda_s^{(i)})^r (r - 1)!$$

Then the relationship between A_r and g_{m_s} is, apart from the constant, of the same form as that between the moment generating function and the cumulant generating function, this permitting the use of standard tables.

2. *Solution for finite t*
After simplification and abbreviating

$$Pr\{N_1(t) = i, N_2(t) = j, N_3(t) = k \mid N_1(0) = i_0, N_2(0) = j_0, N_3(0) = k_0\}$$

to $p_{i,j,k}$ the general differential equation of the model subject to seven boundary conditions, is given by

$$\begin{aligned} p_{i,j,k} = -Sp_{i,j,k} &+ \beta_1 p_{i+1,j,k} + \\ &\alpha_1 p_{i-1,j,k} + \\ &\gamma_{21} p_{i-1,j+1,k} + \\ &\gamma_{12} p_{i+1,j-1,k} + \\ &\gamma_{23} p_{i,j+1,k-1} + \\ &\gamma_{32} p_{i,j-1,k+1} + \\ &\gamma_{31} p_{i-1,j,k+1}. \end{aligned} \quad (14)$$

$$i = 1, 2, \ldots \quad j = 1, 2, \ldots \quad k = 1, 2, \ldots \quad t > 0$$

An expression can now be obtained which involves the 3-variate generating function $G(z_1, z_2, z_3, t)$

$$G(z_1, z_2, z_3, t) = \sum_{i=0}^{i=\infty}\sum_{j=0}^{j=\infty}\sum_{k=0}^{k=\infty} p_{i,j,k}\, z_1^i z_2^j z_3^k$$

$$\begin{aligned} G'(z_1, z_2, z_3, t) = &\left(-S + \frac{\beta_1}{z_1} + \alpha_1 z_1 + \gamma_{21}\frac{z_1}{z_2} + \gamma_{12}\frac{z_2}{z_1} + \gamma_{23}\frac{z_3}{z_2}\right. \\ &\left. + \gamma_{32}\frac{z_2}{z_3} + \gamma_{31}\frac{z_1}{z_3}\right) G(z_1, z_2, z_3, t) \\ &+ \left[(\beta_1 + \gamma_{12}) - \frac{\beta_1}{z_1} - \gamma_{12}\frac{z_2}{z_1}\right] G(0, z_2, z_3, t) \\ &+ \left[(\gamma_{21} + \gamma_{23}) - \gamma_{21}\frac{z_1}{z_2} - \gamma_{23}\frac{z_3}{z_2}\right] G(z_1, 0, z_3, t) \\ &+ \left[(\gamma_{32} + \gamma_{31}) - \gamma_{32}\frac{z_2}{z_3} - \gamma_{31}\frac{z_1}{z_3}\right] G(z_1, z_2, 0, t) \end{aligned} \quad (15)$$

Now using the Laplace transform over the positive real axis

$$G^*(z_1, z_2, z_3, s) = \int_0^\infty G(z_1, z_2, z_3, t) e^{-st} dt$$

$$G^*(0, z_2, z_3, s) = \int_0^\infty G(0, z_2, z_3, t) e^{-st} dt \text{ etc.}$$

Then (15) reduces to

$$G^*(z_1, z_2, z_3, s) = \frac{N}{D}$$

where

$$\begin{aligned} N = &z_1^{i_0} z_2^{j_0} z_3^{k_0} \\ &+ \left(\beta_1 + \gamma_{12} - \frac{\beta_1}{z_1} - \gamma_{12}\frac{z_2}{z_1}\right) G^*(0, z_2, z_3, s) \\ &+ \left(\gamma_{21} + \gamma_{23} - \gamma_{21}\frac{z_1}{z_2} - \gamma_{23}\frac{z_3}{z_2}\right) G^*(z_1, 0, z_3, s) \\ &+ \left(\gamma_{32} + \gamma_{31} - \gamma_{32}\frac{z_2}{z_3} - \gamma_{31}\frac{z_1}{z_3}\right) G^*(z_1, z_2, 0, s) \end{aligned} \quad (16)$$

$$\begin{aligned} D = s + S - \frac{\beta_1}{z_1} - \alpha_1 z_1 - \gamma_{21}\frac{z_1}{z_2} - \gamma_{12}\frac{z_2}{z_1} \\ - \gamma_{23}\frac{z_3}{z_2} - \gamma_{32}\frac{z_2}{z_3} - \gamma_{31}\frac{z_1}{z_3} \end{aligned} \quad (17)$$

The relationships expressed in equations 16 and 17 would in most cases be sufficient to find an explicit solution for the trivariate probability generating function $G(z_1, z_2, z_3, s)$. In this particular instance, however, more information is needed to evaluate $G^*(z_1, z_2, 0, s)$, $G^*(z_1, 0, z_3, s)$ and $G^*(0, z_2, z_3, s)$

A method of attack which could perhaps prove useful if explored in greater depth is the application of spectral theory (Lederman and Reuter, 1954; Karlin and McGregor, 1957a, b; Karlin and McGregor, 1958).

At this stage further investigation for finite t does not seem worthwhile as the data will almost certainly lead to some changes in the flows or boundary conditions which could make the problem more or less tractable.

Prerequisites of model use

The numbers, N_i, in each of the three sets refers to EDP applications. A problem therefore arises of defining applications in such a way that at least it becomes possible to count how many are being applied, at any time, in a user organism. This however is not sufficient as a great deal of heterogeneity almost certainly exists between EDP applications. Therefore, it will most probably be necessary to consider more than one unique α_1 and meaningful differences between applications, expressed quantitatively, become important.

The model refers to a single user organism. This requires clarification as it is difficult both to distinguish between different types of entities and to decide if they are users or non users. Operational definitions are therefore needed for such terms as organisation, establishment and system. Usership is related to possession of various types of EDP equipment and a method must be evolved for its consequent identification and classification.

The flows are functions of economic and business variables linked to a set of user organisms or to one particular member. Hypotheses must be set up as to which variables are related to the phenomena and the necessary data collected. Problems of measurement occur as the model is restricted to quantitative inputs.

These points are discussed in greater detail in the following sections.

EDP application analysis

This section presents the proposed method of application analysis.

Application order and counting

Any EDP application can be thought to have two major facets which may be interrelated or simultaneous in time. The first of these, the most costly, is the process of collecting, manipulating, updating and storing the necessary data whilst the second is the set of tasks actually performed on the data files once these have been set up.

Any user organism can perform a variety of tasks on any one or more sets of data files. It can even perform tasks which require a combination of different sets of data files.

At any point in time the constraint on the nature of the tasks and the stable element in computer activity is therefore the data available and present in the files to which the CPU has access together with that stored in its own memory core. It seems only natural, therefore, to consider applications as crucially related to sets of data on which various tasks are being performed.

Suppose a printout of a set of data, run for several tasks, is prepared, this printout will consist of one or several lists of objects. Objects in any one list will usually have (or can be conceptually assigned) consecutive identification numbers. It is possible in this way to split up the entire data set, utilised by a user organism, into data lists. Now the tasks performed on these lists can also be grouped into sets so that the tasks in any one subset require running the same subset of lists as necessary and sufficient data for their execution.

It is proposed therefore that, *one* application is *counted* for each of the subset of tasks performed on the same subset of lists. To distinguish between applications associated with different numbers of lists the application *order* is the number of lists involved.

Suppose for example the user organism 0 has five lists L_1, $L_2 \ldots\ldots L_5$ and eight tasks $T_1, T_2, \ldots\ldots T_8$.

T_1 and T_2	require only running	L_1
T_3	requires only running	L_2
T_4	,, ,, ,,	L_3
T_5	,, ,, ,,	L_4
T_6	,, ,, ,,	L_5
T_7	,, ,, ,,	L_4 and L_5
T_8	,, ,, ,,	L_1, L_2, L_3, L_4

This user organism has five first order applications, one second order application and one fourth order application.

As a practical example, invented only to clarify these points, it is useful to consider two establishments A and B in two different oil companies. They have the following characteristics:

Establishment A

Tasks performed	*Data lists involved*
Pay	Employees
Quality control	Oil wells
Stock control	Ships
Quality control	Oil storage installations
Linear programming	{ Oil wells, Ships, Oil storage installations

Establishment B

Tasks performed	*Data lists involved*
Quality control	Oil wells
Stock control	Ships
Quality control	Oil storage installations

It can be seen, from this information, that A has four data lists, four first order applications and one third order application whilst B has three data lists and three first order applications. It is worth noting that if the usual application descriptions were used B would have only two applications (quality control and stock control).

This information on its own is of limited value as an application characterisation and the next sections show how it can be supplemented by constructing scales for measuring the similarity between applications and by using classification variables.

Scales for application measurement

Application measurements can refer to either observable properties of the tasks or those of the database employed in their execution. The scales suggested here cover both these entities and furthermore are thought to be relevant to data transmission usage. At this stage thir relevance is, of course, only hypothetical and confirmation can only be obtained when data becomes available. The properties which have been considered for measurement are, as far as possible, directly observable. Reproducibility has been maximised by making measurements as totally free as feasible from respondent attitudes or opinions and from interviewer interpretation. Most of the scales in this section only imply an ordering along a certain dimension. It should be possible to convert these ordinal scales into interval or ratio scales when the necessary application distributions are known. Similar procedures are used in sociological research for dealing with social classes.

1. *Application size*

An application of any order consists of a set of tasks together with a data set which is both necessary and sufficient for their execution. The entities in the data lists which make up the data set are the input units, examples are customers, suppliers, establishments and measurements. The size of an application is the number of input units. This variable is supplemented by the average amount of information, in bits, which is associated with the input units.

These measures are clearly related to the quantity of work undertaken for the application both in data collection and by the EDP equipment.

2. *Application usage*

The objective is to measure the importance of the application in terms of computer time and throughput. Slight differences occur according to whether the application tasks are performed in discrete runs or continuously. Measurements are taken of either the average time/run and the average number of runs/unit time or the average running time/unit time. The average transaction flow/unit time is used in both cases.

3. *Application data access*

This variable characterises an application by the rapidity with which data must be made available to the central processor. The scale has three points which are: instantaneous, direct access (core), fast, indirect access (disc or drum) and slow, indirect access (tape or cards).

4. *Application data permanence*

The time validity of the application data set is described by this scale which has the following four points:

(*a*) *Imaginary*

The information stored is completely fictitious as, for example, in programme testing.

(*b*) *Temporary*

In this case the information is collected for a series of analyses on a 'once off' basis. The result is usually in the form of a report, document etc. which terminates the work. Such applications as market research, some forms of simulation, industrial experiments, most scientific research are included here.

(*c*) *Semi-permanent*

The information on the input units is collected either continuously or updated at fairly regular intervals as standard procedure. Task performance is dependent on, and posterior to, updating. The function of the tasks is usually to produce time dependent information on a dynamic situation, influence it or register changes that have occurred.

(*d*) *Permanent*

The information on the input units is collected either continuously or updated at fairly regular intervals as standard procedure. Task performance and updating are independent. The function of the tasks is usually to provide almost time independent information on a fairly static situation. Information banks fall into this category, good examples are automated land registry search and computerised medical diagnostic systems.

This scale will be used in conjunction with the average update interval (average time which elapses between consecutive updating of the input units).

5. *Application task continuity*

This scale has been designed to show the degree of involvement of the user organism with the application. It has four points which are as follows:

(*a*) *Sporadic*

The application is 'once off' or carried out for fortuitous reasons at very irregular intervals. It is not a constant feature of the user organism work load.

(*b*) *Regular*

The application is carried out at regular intervals and forms a constant feature of the user organism work load.

(*c*) *Semi-continuous*

The application is carried out continuously except for maintenance down time and hours when the user organism is non-functional. An automated technical reference library available weekdays from 9 a.m. to 12 p.m. would be in this category.

(*d*) *Continuous*

The application is carried out continuously except for maintenance down time or breakdown.

6. *Application complexity*

The aim of this scale is to show the fundamental calculation complexity of the application. The three points are as follows:

(*a*) No computation: The output is in the form of lists, no calculations are required although some classification may take place.

(*b*) Simple operations: Calculations are confined to characteristics of individual input units and no descriptive information is produced, in numerical form, on the set of input units.

(*c*) Complex calculations: This includes all applications where the output requires combination of the numerical information stored on several input units. Calculation of basic statistics, simulation, balance sheets, time series analysis etc. are of this type.

7. *Application deployment*

This scale shows how much authority the user organism has delegated to the application. The four points are as follows:

(*a*) No authority: The output of the application does not influence decisionmaking in the organisation at any level.

(*b*) Decision Guidance: The work at this stage can be considered as an aid to decisionmaking; it puts the basic information in a form useful for management perusal. The output is usually in the form of tables, graphs, etc. which are then examined by management; subjective considerations are brought in and a decision is made by the people concerned.

(*c*) Implicit Authority: The output is dynamic; the EDP system takes a decision which is then implemented by human intervention, usually as a matter of routine. Such operations as acceptance and rejection of passengers, credit accounts, ordering of raw materials, etc. are all included here. Middle management can only rectify decisions, in special circumstances, after they have been taken and usually implemented.

(*d*) Explicit Authority: This is similar to (*c*) except that the EDP system actually implements the decision without human intervention. Good examples are process control, regulation of water levels in dams, etc.

Application classification

The variables described in this section consist of application input units, their data characteristics and output units. These classification criteria, in conjunction with the application measurement scales, are thought to be of considerable value in providing explanatory factors, when analysing different types of data transmission usage patterns and volumes.

1. *Input units*

The input unit classification consists of ten categories. Six subcategories are used to distinguish between different types of processes. Definitions are as follows:

(*a*) *Customers*
These are the entities, either persons or organisations, for whom the articles are manufactured, the natural resources extracted, the agricultural goods farmed or generally the work done by the application source, establishment or organisation. They can be private purchasers, ratepayers the general public, etc.

(*b*) *Suppliers*
These are the entities, either persons or organisations, for whom the application source, establishment or organisation acts as customer.

(*c*) *Process*
This is a group of functions carried out by the application source establishment or organisation. Examples are production, extraction, maintenance, repair, research, etc. Several different types of process can be distinguished and are summarised below:

(i) Transformation
The vast majority of industry falls into this category. It entails physical work done on matter which either removes it from its natural environment, changes its quality or appearance or alters some other one of its physical characteristics. Included here are: Farming, stock-rearing, market gardening, ploughing, crop-spraying, fishing, planting and re-planting of forests, processing, packaging, cleaning, construction, catering and generally every type of manufacture, maintenance, repair, installation and extraction.

(ii) Transportation
This is work performed on matter, people or information whose primary objective is to change the spatial co-ordinates. Included here are: Air, sea, land transport of people and merchandise, transport of information over air, wire or land, etc.

(iii) Dealing
This is a purely commercial process involving only buying, selling, leasing or renting by a source organisation or establishment either on its own account or for another organisation or establishment. Included here are: Agents, dealers, warehousing, retailing, wholesaling, leasing, letting, etc.

(iv) Investment
This process is simply to facilitate the transmission of funds in space or time. Included here are: Banking, insurance, mortgages, all types of money lending, etc.

(v) Advice
This process is the elaboration, combination and analysis of information. Included here are: Exploration, research, design, development, consultancy, professional services not included in the following paragraph.

(vi) Miscellaneous Services
These are people-oriented services usually performed by central or local Government authorities, institutions, etc. This category contains exclusively: Educational services, medical and dental institutions, religious organisations, cinemas, radio, theatre, welfare and charitable services, trade and professional institutions and those functions of defence, police and fire, not included in other paragraphs.

(*d*) *Process input*
These are the entities obtained from suppliers, the natural environment or possibly customers on which the process is performed.

(*e*) *Process output*
These are the resultant entities, once the process has been performed which are supplied to customers.

(*f*) *Equipment*
These are all the machines present in the organisation and which in some way, either directly or indirectly, have a bearing on the process carried out.

(*g*) *Personnel*
These are the people employed by the organisation, either as permanent staff or casual labour and who cannot be considered suppliers of it (i.e. not consultants, etc.).

(*h*) *Establishments*
Those premises, used by a source organisation or source establishment, either directly or indirectly in order to perform the process.

(*i*) *Financial input*
These are the shareholders of the source organisation.

(*j*) *Financial output*
These are the other organisations of which the source organisation is a shareholder.

2. *Data characteristics*
Each of the ten input unit categories is divided into subcategories. These distinguish between different variables which are used, in the application, as input unit characterisations. A full list is given in Appendix 3.

3. *Output*
Output units are only defined for applications which provide decision guidance or have implicit or explicit authority. These units are then those entities directly affected by the application.

The classification for the output units consists of exactly the same ten categories as devised for the input units. In many applications input and output units will coincide, in payroll for example, both types of units would obviously be 'employees'. An illustration of when these units differ would be a production organism which sets up a process scheduling application. The input units could be customers with process output requirements as data characteristics whilst output units could be 'process output'.

Examples of application description
Four purely hypothetical examples are described and then coded (**Table 1**) to highlight points of similarity and difference. The examples are scientific, payroll, linear programming and process control applications.

Although only the basic application coding will be given here, the questionnaire that is being developed to obtain the application descriptions also requests details of the geographic flows of data and the types of establishments involved.

1. *Scientific*
Scientific analysis of a problem in genetics is performed by inputting cards inhouse to a large computer for batch processing. The printout is received directly.

2. *Payroll*
Payroll information is transmitted in punched card form from five warehouse locations to a central headquarters containing the CPU. After processing, the printout of pay details is sent back to the remote locations for local payment of employees.

3. *Linear programming*
Profit optimisation analysis is carried out on a large computer, involving regulation of oil production at various sites, the oil storage and its distribution. Linear programming is employed.

4. *Process control*
Automatic control of a large continuous process plant, for example an oil refinery.

Table 1 shows an increase of overall sophistication of applic-

Table 1

CHARACTERISTIC	SCIENTIFIC	PAYROLL	LINEAR PROGRAMMING	PROCESS CONTROL
1 Application order	First order	First order	Third order	First order
2 Application size				
Input units	3,600	750	200	350
Bits/input unit	240	800	2,480	80
3 Application usage				
Discrete				
time/run (hrs)	1	4	8	—
input/run (bits)	800,000	32,000	160,000	—
output/run (bits)	8,000	304,000	160,000	—
runs/month	1	8	20	—
Continuous				
hrs/month	—	—	—	672
input (bits/month)	—	—	—	$0{\cdot}9 \times 10^{11}$
output (bits/month)	—	—	—	$1{\cdot}2 \times 10^{11}$
Transaction flow (million bits/month)	0·8	2·7	6·4	210,000
4 Application access	Slow	Slow	Slow	Instantaneous
5 Application data permanence	Temporary	Semi-permanent	Semi-permanent	Semi-permanent
6 Application continuity	Sporadic	Regular	Regular	Continuous
7 Application complexity	Complex calculations	Simple operations	Complex calculations	Complex calculations
8 Application deployment	Decision guidance	Implicit authority	Implicit authority	Explicit authority

ation in moving from left to right. The attributes of deployment, continuity, data permanence and usage all increase when moving in this direction. This accords with the actual historical development of these applications—a reason for belief in the utility of the classification. Also shown in the table are those relationships which are similar for quite differently ranked applications, e.g. application complexity of both the 'scientific' and the 'process control' applications and the similar size of the databases of the first three applications. For the purposes of analysis different linear, or non-linear weighted combinations of these scales can be considered. These weights can be either objectively derived, or hypothesised, and afterwards their relative merits tested for the purposes of the analysis.

Terminology

One of the defects in most data transmission research is the lack of precise definition in the terms used. This leads to incompatibility between the results of different studies and to difficulties of interpretation. Rigorous definitions of various items of EDP equipment, user organisms and traffic volume units are fully set out in Appendix 1.

In many instances current terminology has been avoided. Often terms in general use cover a variety of concepts and confusion occurs when their normal meaning is restricted or slightly altered by redefinition. It seems preferable, in these situations, to create a new nomenclature.

The definition of a piece of EDP equipment has been made to depend on its characteristics, on the work it performs and on how it is placed in the network.

The user organism definitions have been made to match, as closely as possible, those found in official statistics as this is convenient for data collection purposes and subsequent analysis.

Some of the terminology is specifically linked to applications. Measurement and classification of applications is based on observing a set of tasks combined with a set of necessary and sufficient data. It is important to realise that these descriptions are not invariant under change of observer location. The most obvious example is an organisation which has constituted some database D. Suppose that two establishments A and B use this base as necessary and sufficient data for task sets T_A and T_B respectively. The application description will then refer to (D, T_A) or (D, T_B) according to whether it is observed from establishment A or B. Furthermore the description generated by considering the organisation itself, rather than the member establishments, as the user organism would be based on $(D, T_A \cup T_B)$. Measurements of task continuity, of complexity and of deployment as well as the nature of the input units can all obviously be affected.

Observations on data collection and analysis

Aspects of data collection

The success of this project depends entirely on the willingness of computer users to provide the Post Office with information. It is necessary that a clear statement of the aims of the data collection activity and its subsequent analysis is given both to intended respondents and the public as a whole. Pilot surveys would have to be made to determine the feasibility of the whole project. Abstracts of the results of the general analysis should also be made available.

The existence of a UK file of computer installations maintained by the National Computing Centre indicates that the approach most likely to be effective for data gathering lies in starting the investigation at the computer centres and then proceeding to all the associated user organisms. Data about a user organism may in many cases have to be acquired from a number of different respondents. It would be an aim of the study to maintain very close contacts between interviewer, respondent, and research organisation. Finally and most importantly the preservation of confidentiality of sensitive information would be a paramount objective.

EDP data

The general EDP oriented data which will be placed on file, if the methodology outlined in the first two sections of this paper proves feasible, is given in Appendix 2. Care has been taken to restrict collection to opinion-free accessible information so as to reduce to a minimum the work load placed on voluntary respondent user organisms.

The section on CPU data covers the same ground as the

National Computer Index. The data file proposed here is in fact a natural extension of the National Computer Index which is now in operation.

Business, survey and economic data

The volume of data transmission generated by a user organism could be closely related to certain business and economic variables. This possibility has in fact been explicitly incorporated into the model discussed in a previous section. The type of data should, as far as possible, be restricted to and similar in format with that collected for other official statistics. This approach is useful because the user organisms who decide to co-operate have already had to prepare these returns and it also permits comparison with published material. Some of the data collected for the Censuses of Employment, Production, Retail and Distribution and for the Survey of Research Expenditure seems also to be very relevant in the data transmission field.

Nevertheless information of this type, for complex phenomena such as data transmission, can only be useful if it is at a sufficient level of detail. The number of employees in a user organism, for example, is most probably far too general. Much more informative would be, however, the numbers of employees in a user organism in various categories (administrative and clerical, scientific, technical, operatives).

Geographic data

It is also likely that the geographic spread of a user organism plays an important role in data transmission. Although it is most probably fairly easy to acquire location data on the establishments which form the more complex types of user organisms, effective and realistic measure construction gives rise to a number of non-trivial problems.

As a first step it is possible to consider the isocentre and dispersion of a user organism or a set of user organisms. These concepts and their many useful theoretical properties are fully discussed in (Vo-Khac Khoan, 1965; 1967).

In outline suppose a complex user organism (system, organisation) consists of n establishments. Let the vector $\vec{x}_i$ characterise the geographic co-ordinates (a_i, b_i) of each establishment and the scalar w_i some quantities of interest such as data traffic volume or number of employees. Then the simple isocentre I, the weighted isocentre I_w, the simple dispersion D and the weighted dispersion D_w are defined by the following expressions.

$$\vec{I} = (\sum_{i=1}^{i=n} \vec{x}_i)(n^{-1})$$

$$\vec{I}_w = \sum_{i=1}^{i=n} w_i \vec{x}_i$$

$$D = \sum_{i=1}^{i=n} \|\vec{x}_i - \vec{I}\|$$

$$D_w = \sum_{i=1}^{i=n} \|\vec{x}_i - \vec{I}_w\| w_i$$

Where $\sum w_i = 1$ and $\|\vec{x} - \vec{y}\|$ is the square of the distance (Euclidean metric) between two points with vector co-ordinates $\vec{x}$ and $\vec{y}$.

The isocentre and dispersion of a system or organisation depend only on the co-ordinates of the establishments whilst network measures would require knowledge of how these are linked up. If, therefore, relationships could be found between the weighted or unweighted:

(*a*) isocentre and the command centre (CHQ, CPU)

(*b*) dispersion and the network square distance measures

advantages would accrue from modelling the real situation by simpler, and hence easier to forecast, quantities.

Incidentally the isocentre has the property that the total square distance between it and the establishments in the system or organisation is a minimum. This is therefore the optimum location if cost of transmission depended solely on distance. This also holds with the weighted dispersion and weighted isocentre which is optimum if costs depend on distance weighted by relative quantity transmitted. The isocentre suffers the disadvantage that it is not invariant when element i is added or subtracted whereas the command centre usually is.

Some possible interpretations and uses of these measures are:

(*a*) The Isocentres

They show the 'centres of communication' for systems and organisations. In any set of such entities the isocentre of either the real centres or isocentres could be useful for the establishment of lines, switching centres, etc.

Furthermore, if total system or organisation transaction volumes are associated with the isocentres, a communication load density distribution over the country can be obtained.

(*b*) Average dispersion/establishment

This is a measure of the isolation of establishments in an organisation or system.

(*c*) The ratio: load/dispersion

Here load is the EsVT of the system or organisation and this ratio might then be related to the economics of data transmission installation.

(*d*) Dispersions—weighted and unweighted

Perhaps a relationship exists between dispersion and probability of DT transmission.

Conclusions

This paper has described the development, in recent years, of Post Office thinking about data transmission user research. Part 1 dealt with the overall picture whilst Part 2 has concentrated on a detailed exposition of the more important technical facets believed to be of general interest. At the moment no commitment exists for the research methodology presented in Part 2. In fact a major objective of this publication is to receive, before any specific programme is set in motion, suggestions, criticisms and comments which readers might care to make.

Acknowledgements

The authors wish to thank Mr. J. H. Hayter (Statistics and Business Research Department, The Post Office) for his many useful contributions, and senior officials in Telecommunications Headquarters, Data Communication Marketing Division, The Post Office, for their interest in this project.

Appendix 1 Terminology

1. *EDP equipment*

(*a*) Central Processing Unit (CPU)

A piece of EDP hardware which is programmable, has immediate storage facilities, processes data, performs calculations and controls the system by conducting the 'highest level' work.

(*b*) Terminal-end equipment (TEE)

A unit (or part of a unit) of peripheral equipment which provides the contact between the system and the environment.

TEEs can be divided into two categories:

(i) TEE 1

The TEE are in a different establishment from the CPU and connected to it by public or leased lines possibly passing through some intermediate equipment or other TEE.

(ii) TEE 2

The TEE are in the same establishment as the CPU and connected to it by private line through, possibly, some intermediate equipment, also in the same establishment.

(*c*) Communication interface (CI)

A device which permits data signals to pass on to the public or leased line communication network. It provides the link between TEE, CPU or intermediate equipment and the leased or public line.

(*d*) Intermediate equipment (IE)
A unit (or part of a unit) which processes information and in this way facilitates the CPU operations of computing, classification and information storage and line control.

Four categories of IE can be defined:

(i) IE 1
The CPU, IE and TEE are in the same establishment and linked by private wire, possibly by way of other IE.

(ii) IE 2
The CPU and IE are in the same establishment and linked by way of private wire whilst the TEE is in another establishment and connected to the IE by public or leased line, all, possibly by way of other IE.

(iii) IE 3
The IE and TEE are in the same establishment and linked by private wire whilst the CPU is in another establishment and connected to the IE by public or leased line, all, possibly by way of other IE.

(iv) IE 4
The IE, TEE and CPU are all in different establishments and linked by public or leased line possibly by way of other IE.

(*e*) Batch equipment (BE)
A device used for the preparation of output data or modification of input data but not telecommunication linked to the CPU or IE. Data is transferred to or collected from the CPU by some physical means of transportation.

(*f*) Concentrator (C)
A piece of equipment which arranges data solely to facilitate its flow through the telecommunication network. It is linked to the CPU(s) and TEE(s) by private, public or leased lines both at input and output.

(*g*) Network linkages (NL)
Lines of communication which provide system interconnection. Two categories can be defined:

(i) NL 1
These link an establishment where BE is held to another establishment containing EDP equipment. They are not telecommunication linkages.

(ii) NL 2
These are telecommunication linkages between one establishment and another, both containing EDP equipment.

2. *User organisms*

(*a*) Establishment (E)
An establishment is normally the whole of the premises at a particular address and refers, for example, to a mine, shop, farm or factory. If at some address there are two or more departments (plants, warehouses, etc.) having separate records and different activities then they will be considered as separate establishments.

Four types of establishment can be considered

(i) E 1
This is when a piece or pieces of EDP equipment are present on the premises and the establishment uses at least one of them for the exercise of its activities.

(ii) E 2
This is when a piece or pieces of EDP equipment are present on the premises but the establishment uses none of them for the exercise of its activities.

(iii) E 3
This is when the establishment has no EDP equipment on the premises but uses some other EDP equipment in another establishment for the exercise of its activities.

(iv) E 4
The establishment neither has nor uses EDP equipment.

For any particular application an establishment can be a member of one or more of the following categories:

(i) Application Input Establishment (AIE)
Those establishments with TEE(s) or BE(s) where the data for a particular application is input.

(ii) Application throughput establishment (ATE)
Those establishments with IE(s) or CPU(s) where the data for a particular application is processed.

(iii) Application output establishment (AOE)
Those establishments with TEE(s) or BE(s) where the processed data for a particular application is output.

(iv) Application user establishment (AUE)
Those establishments where a particular application directly provides decision guidance or where it has implicit or explicit authority.

(*b*) Organisation (0)
Different definitions are used according to whether the entity is engaged in commercial (01) or public (02) activity.

(i) 01
If an establishment is registered at Company House or is a sole trader or partnership then normally it is taken as one single organisation. The only exception to this rule is that if two or more establishments are both registered at Company House or belong to the same sole trader, partnership or group (Holding Company) and with very close integration between them (combined central purchasing, personnel, accounting, etc.) and in the same SIC activity, then these will be considered as one organisation together with the HQ, if this exists, where the integration takes place. Should an establishment(s) not be registered at Company House nor be in business and/or belong to a sole trader or partnership then the organisation is that entity registered at Company House which controls their activity directly. This does not refer exclusively to financial control but rather working control. Pure holding companies should be avoided as organisational units.

(ii) 02

Type	*Organisation*
Central Government establishments	Department
Local Authority establishments	Local Authority
Independent Bodies	The Independent Bodies e.g. research council, water board, regional hospital board, university, public schools etc.
Nationalised Industry establishments	Each Nationalised Industry will be considered separately.

An application user organisation (AUO), for a particular application, is that organisation which incorporates one or more of the pertinent AUE(s).

(*c*) System (S)
The system is that group of EDP equipment linked with transaction flows by either NL 1 or NL 2.

In each system there must be at least one CPU or, if in the network no such hardware is present, at the very least a TEE together with some IE which function as a CPU. If in a system there are two or more CPU then each will be said to generate a separate system unless the CPU are directly linked together. Only in this case will such a conglomerate be referred to as a single system.

An application system (AS) is the conglomerate of pertinent AIE(s), ATE(s), AOE(s) and AUE(s).

3. *Traffic volumes*

(*a*) Establishment Volume Input (Es VI)

Quantity of information, in bits, received by the establishment in unit time, through the linkages connected to it.

(*b*) Establishment Volume Output (Es VO)

Quantity of information, in bits, transmitted by the establishment, in unit time, through the linkages connected to it.

(*c*) Equipment Volume Input (Eq VI)

Quantity of information, in bits, received by a piece of EDP equipment in unit time by way of linkages extended to it inside the establishment.

(*d*) Equipment Volume Output (Eq VO)

Quantity of information, in bits, transmitted by a piece of EDP equipment, in unit time, by way of linkages extended to it inside the establishment.

(*e*) Establishment Volume Transaction (Es VT)

The sum, for any particular establishment, of Es VI and Es VO,

(*f*) Equipment Volume Transaction (Eq VT)

The sum, for a particular piece of EDP equipment, of Eq VI and Eq VO.

Appendix 2 Data Collection

System

International linkages.

Organisation

Location, SIC. EDP Activity (Bureau, others).

Establishment

Location, SIC, Function (Regional Office, Branch, etc.), Type (E1, E2, E3, E4), EDP staffing, Traffic Volume, Average call duration, EDP Description (Spur to spur, centre of star, etc.).

Central processing unit

Manufacturer, model, core size, peripheral equipment details, language details, mode of operation (batch, multiprogramming, etc.), remote computing mode (remote job entry, time sharing, etc.).

Terminal end equipment

Description (VDU, lineprinter etc.), type (TEE1, TEE2) manufacturer, model number.

Communications interface

Ownership, datel service code, manufacturer, speed, tariff code.

Intermediate equipment

Manufacturer, model, type (IE1, IE2, IE3, IE4).

Batch equipment

Description (lineprinter, card punch, etc.).

Concentrator

Manufacturer, model, local line speed, aggregate signal speed.

Network linkages

Description (public, leased, courier), length, traffic volume, speed.

Application

Application user establishments. Application input establish ments, application output establishments, application throughput establishments, application user establishment description, application user organisation, application user organisation description, application system description, application traffic volumes.

A reference system will permit analysis of relationships existing between the user organisms, their EDP equipment and their applications.

Appendix 3 Data characteristics

Customers

Process output requirements (real or potential), process output received financial situation, other characteristics useful for their classification.

Suppliers

Process input requirements, process input received, financial situation, other characteristics useful for their classification.

Process

Measurement on the process itself (indirect measurements) and which serve as a guide either in terms of the speed or quality of the resultant process output.

Measurement on the matter moving through the process, which serves as a guide either in terms of the speed or quality of the resultant process output.

Process input

Measurements on the quality of the entities, measurements on the rate and quantity of entry.

Process output

Measurements on the quality of the entities, measurements on the rate and quantity of exit.

Equipment

Measurements on the characteristics of the equipment (such as age, rate of function, time since maintenance), financial situation.

Financial input

Identification and Financial Situation of shareholders, other characteristics useful in their classification.

Financial output

Identification of investments, other characteristics useful in their classification.

References

COX, D. R. and MILLER, H. D. (1965). *The Theory of Stochastic Processes*. Methuen, London.

FELLER, W. (1965). *An Introduction to Probability Theory and its Applications*. Wiley, New York.

KINGMAN, J. F. C. (1969). Markov Population Processes, *J. Appl. Prob.*, Vol. 6, pp. 1-18.

WHITTLE, P. (1967). Nonlinear Migration Processes, *Proc. 36th Session of the Int. Stat. Ins.*

WHITTLE, P. (1968). Equilibrium distributions for open migration processes, *J. Appl. Prob.*, Vol. 5, pp. 567-571.

BAILEY, N. T. J. (1954). A continuous time treatment of a simple queue using generating functions, *J.R Statist Soc.*, Vol. *B* 16 288-291.

LEDERMANN, W., and REUTER, G. E. H. (1954). Spectral theory for the differential equations of simple birth and death processes, *Phil. Trans. Roy. Soc.*, London Vol. 246, pp. 321-369.

KARLIN, S. and MCGREGOR, J. L. (1957b). The Differential Equations of Birth and Death Processes, and the Stieltjes Moment Problem, *Trans. Amer. Math. Soc.*, Vol. 85, pp. 489-546.

KARLIN, S., and MCGREGOR, J. L. (1958). Many server queuing processes with Poisson input and exponential service times, *Pacific J. Math.* Vol. 8, pp. 87-118.

KARLIN, S. and MCGREGOR, J. L. (1957a). The classification of birth and death processes, *Trans. Amer. Math. Soc.*, Vol. 86, pp. 366-400.

VO-KHAC KHOAN (1965). *Etude Théorique du Problème de Segmentation*, CEGOS, Paris.

VO-KHAC KHOAN (1967). *Aspect Théorique et Aspect Pratique de la Segmentation par la Méthode des Moindres Carrés*, CEGOS, Paris.