

INFORMATION STRUCTURES

Here are some notes on the structure of information managed by the Computer Science department. They are exploratory in nature, and intended more as a means of presenting examples than as a definitive document.

If I am to coordinate the information flows through the Computer Science department, I have to find out a number of things. Here are some examples :

- 1 : What do I mean by "coordination" ?
- 2 : What is the information which needs coordinating ?
- 3 : How does the information move around ?
- 4 : What sort of strategy is appropriate for coordinating it ?

I'll address these points one by one, and doubtless get sidetracked onto other more or less related points as I go.

WHAT IS COORDINATION ?

I shall begin with a working definition which seems to be right before the event. I am quite ready to accept that the definition will have to be changed in the light of experience, but I can't do that until I get the experience, and I have to start somewhere. I therefore propose this definition :

A body of information is coordinated when it has these properties :

1 :	It exists in some identifiable form;	If the information is worth having, it should be recorded somewhere. Relying on someone or other to remember it or reinvent it when it's needed isn't satisfactory.
2 :	No two items conflict;	If the information makes sense, it isn't contradictory. This statement might not be strong enough - it doesn't allow for contradictions which depend on more than two items ($A > B$; $B > C$; $C > A$). If we need it stronger, we can make it stronger, but I'm not sure that I see how to check it easily.
3 :	All dependencies between items are known;	- otherwise conflicts can arise unbeknownst. (So perhaps this is included in property 2 ? Never mind - it's important enough to emphasise.)
4 :	It is up to date;	New information must be conveyed quickly through the system to all relevant parts - and old information must be checked from time to time to make sure that it hasn't dropped behind.
5 :	Procedures for managing change are defined and operated;	That includes all sorts of change : changes in information from outside, changes in people's circumstances or functions, changes in what we do (new things start, old things stop).
6 :	Any constraints on the use of information are identified and implemented.	This covers security, confidentiality, etc. I shall absolutely not be rushing into handling confidential information, but the issue should be recognised from the start.

- and, doubtless, much, much more, but that will do to be going on with.

WHAT INFORMATION IS COVERED ?

My rough definition of the scope of this exercise in information coordination is something like "Any information of concern in running the department which is communicated between independent entities (so far that includes people and software)".

It is not intended to include people's private information, nor to poke into the way they manage their responsibilities to the department. Those are black boxes. The private information black box is none of my business. The department responsibility black boxes are of interest in that they require information from time to time, and are usually expected to produce information from time to time, but provided that they do that I have no particular interest in how it's done. It is not impossible that experience will suggest better ways of doing things which might impinge on some such black boxes (no, this is not a veiled threat - I have none in mind as yet); should the event arise, a period of diplomatic negotiation will ensue.

WHAT ARE THE INFORMATION FLOWS ?

I don't know. I don't think that anyone else knows, either. Various people know bits of it, but - judging by the elderly nature of some of the information on display by various means - there do seem to be bits here and there that have been forgotten. I don't expect that it will be very easy to find out just what determines what, and how, so don't expect any sudden revelations for a while.

What is clear is that there's a lot of information about, and it flows in many strange ways. There are a few local areas where the flows are recognised and already organised - the office and the booking system come to mind, and now I've offended everyone else with a system but it's just for the sake of an example - but there are a lot of muddy bits in between. I found some of them when administering examinations and diplomas.

In general terms, information comes in, stands about for a while in various repositories, moving a bit from time to time, then (perhaps) goes out. Input comes from the registry, faculty, school, our own activities, and other sources; output goes back to the same places, and also to students, and to the world at large. We have various collections of information, some maintained manually and some by computer means of various sorts. Information flows around in overt and covert ways. I have a touching belief that if we find out just what we need to do and what's there, we should be able to do it better.

One further important characteristic of this network of information is that it keeps changing. That's not just the information; the network itself keeps changing. New regulations come along, we do things in new ways, and so on. A rigid system isn't likely to last for long, so flexibility is an important property.

HOW CAN IT BE MANAGED ?

The system is a rather complicated (though not obviously very complex) set of interacting processes executed by a distributed system of processors (people and machines) communicating through unreliable channels. (That's perhaps a more charitable way of saying that some processors don't always respond very quickly to prompts for action.) It could therefore perhaps be handled by means appropriate to database management or communications.

As well as being concerned with communicating data, though, it is a real-time system. Some of the information must be produced to deadlines, and the requirement for consistency implies that new information should be disseminated quickly (not in itself a real-time requirement, but in the same area). As I also know more about real-time systems, that's the way I'll approach it in the first instance.

The first job is therefore to write down a system specification. I propose to follow an oldish but effective technique which begins by identifying outputs and works backwards therefrom to discover the communication structures. (In case that sounds a bit rarefied to those eager to see concrete results, I remark that it's a sensible place to begin anyway; it identifies the outputs and who's responsible for them, which sounds to me like a promising place to start by any criterion.)

I'm aware that various people engaged in running complex information systems have probably worked out all manner of ways of managing systems of this type. I'm proposing to start from scratch not simply on the "not-invented-here" principle, but on the grounds that I haven't time to take a course on organisational management, or whatever, and that the method I know already has a good chance of working well enough to be useful.

EXAMPLES.

1 : NOTICE BOARDS.

There are several notice boards around the department, and at least one about the department but not really in it - in the ground floor foyer. They are perhaps not our most significant information channels, but they're easy to think about and therefore a good example for this purpose.

They all contain different sorts of information; I don't think any two are precisely identical, but I haven't compared them carefully yet. Some have people's names, some have their functions, some have their photographs, and so on. Some change to some notice board somewhere might therefore be required whenever someone arrives or leaves, whenever someone's function changes, etc.

Therefore, whenever one of these events occurs, someone - the notice-board-manager (NBM), for want of a better term - should be notified. Who is the NBM ? How should the notification happen ?

Evidence suggests that at present Anita and Neena change the notice boards from time to time. Does this mean that Anita and Neena are NBMs ? No; it means that the secretarial staff are NBMs, which isn't quite the same thing. There are suggestions of a non-trivial data structure.

2 : HANDBOOKS.

The handbooks are examples of information channels with distributed responsibility. Someone is in charge of the overall process, but many other people are responsible for parts of the content, and quite a lot of it is linked to other sources of information - we need lists of our lecture courses for all manner of purposes, for example, and having the handbook available via the WWW introduces further dependencies.

How do we get the handbooks ready on time ? (Unreliably, if the past is any guide, but I'm supposed to be doing something about that. Maybe next year ...) People have to be made aware, somehow or other, that it's time to think about handbook entries; and the job should be made as easy as possible. At the same time, there are deadlines of a sort, and they have to be taken into account.

3 : TWO SITES.

The split-which-sometimes-isn't-a-split (depending on what's convenient at the time) isn't going to make life much easier. At present, both Tamaki and city sites have notice boards and handbook entries, and doubtless those at least will continue. Geography guarantees that there will always be some differences so long as we have a split department, but I have no idea what to do about it, because I don't yet know what it is about which something has to be done. I mention the topic just to show that I'm conscious of it as an additional dimension of complexity.